A STUDY OF THE PERFORMANCE OF LINEAR AND NONLINEAR FILTERS

VIDAL R. ALGAZI

TECHNICAL REPORT 420

JUNE 22, 1964

UNPUBLISHED PRELIMINARY DATA

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

A STUDY OF THE PERFORMANCE OF LINEAR
AND NONLINEAR FILTERS

Vidal R. Algazi

Abstract

29198

We are concerned with the filtering of a message from additive noise when message and noise are statistically independent processes, and our purpose is to gain some insight into the process of filtering. To this end we investigate the performance and behavior of linear and nonlinear no-memory filters.

A measure of the performance of filters is proposed which compares the filter under study to an optimum attenuator acting on the same input. The noise level (noise-to-message power ratio) is emphasized as an independent parameter important to the behavior and the performance of both linear and nonlinear filters. For optimum and nonoptimum linear filters, the effect of the noise level on the performance is the specific object of study.

For nonlinear no-memory filters, after considering the determination of the optimum filter for mean-square and non mean-square criteria, we investigate the characteristics of the message and the noise for which the optimum mean-square filter reduces to an attenuator. As expected, Gaussian message and Gaussian noise will give such a filter but there are many other cases, some critically dependent on the noise level, for which this result occurs.

At the other extreme of behavior, perfect separation by a nonlinear no-memory filter will not occur for any message of interest, and we consider the selection of the message of a defined class which will give the smallest (or largest) average weighted error.

For a given noise, a given error criterion, and a known filter we consider messages that have prescribed peak and average power. We find that a message quantized at four different levels will give, by proper choice of the quantization levels, both the maximum and minimum achievable average weighted error.

The same type of quantized message will be optimum if we now carry out optimization among the filters, as well as among the messages. This result allows us to determine lower bounds on the average weighted error for mean-square and non mean-square error criteria.

TABLE OF CONTENTS

CONTENTS

CONTENTS

v

I. INTRODUCTION

In this work we are concerned with the filtering of a message from additive noise when message and noise are statistically independent processes. In most of the recent work[1-4] one or both of the following aspects of the problem are emphasized: the characterization of the filter belonging to some class and the determination of the filter of this class which is optimum in some sense. Our purpose here is slightly different, and with the ultimate goal of gaining some insight into the process of filtering we study the performance of optimum and nonoptimum filters. We note that for a given noise and a specific error criterion the message statistics determine formally the optimum filter and the resulting performance. Whether or not the performance is satisfactory will depend upon the message statistics, and we would like to know the characteristics of the message which lead to a good or poor separation from the noise. We study specifically linear filters and nonlinear no-memory filters.

1.1 MEASURE OF THE FILTER PERFORMANCE

If we refer, for instance, to mean-square filtering, the success of the filtering operation, that is, how close the output of the filter $y(t)$ is to the message $m(t)$ in the mean-square sense, is measured by the normalized error

$$\frac{\overline{e^2}}{\overline{m^2}} = \frac{\overline{[m(t)-y(t)]^2}}{\overline{m^2(t)}}.$$

The normalized error varies from zero for no error to one for no output, which is the largest error that an optimum filter will give. It can be larger than one for a nonoptimum filter. The normalized error, however, does not indicate specifically the contribution of the filter to the result and, therefore, is not a good measure of filter performance. One way of characterizing the filter performance might be to compare the normalized error at the output of the filter to the "normalized error at the input," $\frac{\overline{n^2}}{\overline{m^2}}$. However, an attenuator, which performs a generally unimportant operation on the input, would have an appreciable performance by this criterion. Here, we shall use the optimum attenuator as a reference system in the measure of the performance of filters. This is, as we shall see, related to the commonly used signal-to-noise ratio performance. Before proceeding with this discussion we shall examine more closely the mean-square performance of the optimum attenuator.

1.2 THE OPTIMUM ATTENUATOR

The output of an attenuator is

$$y(t) = a[m(t)+n(t)]$$

when the input is

$$x(t) = m(t) + n(t)$$

and a is the attenuation constant. The mean-square error is

$$\overline{e^2} = \overline{\{a[m(t)+n(t)]-s(t)\}^2}.$$

If $m(t)$ and $n(t)$ are statistically independent and, furthermore, have zero mean, then $\overline{e^2}$ is minimized for

$$a = \frac{\overline{m^2(t)}}{\overline{m^2(t)} + \overline{n^2(t)}}.$$

Let us define k by

$$\overline{n^2(t)} = k\,\overline{m^2(t)};$$

k is the normalized error at the input which we call the relative noise level. In terms of the power spectra, the factor k is defined by

$$\int \Phi_{nn}(\omega)\, d\omega = k \int \Phi_{mm}(\omega)\, d\omega,$$

in which $\Phi_{nn}(\omega)$ is the noise spectrum and $\Phi_{mm}(\omega)$ the message spectrum.

The expression for the optimum attenuator becomes

$$a = \frac{1}{1 + k}.$$

The normalized error for the optimum attenuator becomes

$$\frac{\overline{e^2}}{\overline{m^2}} = \frac{k}{1 + k}.$$

We see that the optimum attenuator gives an error that is smaller than the error at the input, k. Let us justify this result by physical reasoning. The error at the output of an attenuator

$$e(t) = (1-a)\, m(t) + a\, n(t)$$

has two components, the error resulting from the attenuation of the message $(1-a)\, m(t)$ and the error resulting from noise $a\, n(t)$. (Under the assumption of statistical independence we add the mean-square error resulting from each component to obtain the total mean-square error.) Thus the normalized mean-square error is

$$\frac{\overline{e^2}}{\overline{m^2}} = (1-a)^2 + a^2 k.$$
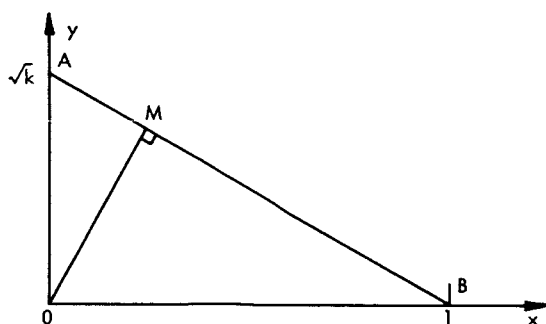
Fig. 1.   Geometric interpretation of the mean-square error
resulting from attenuation.

This can be interpreted as the square of the distance to the origin of a point of rectangular coordinates (see Fig. 1)

$$x_1 = 1 - a$$

$$y_1 = a\sqrt{k},$$

where x_1 is the rms normalized error resulting from message attenuation and y_1 is the rms normalized error resulting from noise. If a is changed with k held constant, the point (x_1, y_1) moves on the line $y = \sqrt{k}\,(1-x)$ in the x, y plane.

The minimum distance to the origin occurs for point M, and the distance is easily found by writing the area of the triangle

$$\text{Area} = \frac{\text{dist(OM)} \times \sqrt{1+k}}{2} = \frac{1 \times \sqrt{k}}{2}$$

$$[\text{dist(OM)}]^2 = \frac{k}{1+k}.$$

The input rms error is represented by the distance OA. The attenuator allows us to choose any point on the segment AB. If the relative noise level k is very large, we get the point B. The minimum distance OM will always be less than or equal to one.

1.3 PERFORMANCE INDEX AND NOISE-TO-SIGNAL RATIO PERFORMANCE

For mean-square filtering we shall use the coefficient

$$\eta = \frac{\overline{e^2}\big|_{\text{opt filt}}}{\overline{e^2}\big|_{\text{opt att}}} = \frac{\overline{e^2/m^2}\big|_{\text{opt filt}}}{\overline{e^2/m^2}\big|_{\text{opt att}}}$$

to measure the performance of an optimum filter. Lubbock,[3] who introduced this coefficient, called it the performance index for noise filters.

Optimum filtering is generally used in two types of applications, both of which lead

3

to different requirements on the filter and to slightly different evaluations of their performance.

In control applications it is generally required that the desired message be approximated as closely as possible at the output of the filter, both in magnitude and in phase. Therefore, any operation that reduces the mean-square error below the noise power at the input will be beneficial. The performance index indicates how a possibly complicated optimum filter ratio compares with a simple attenuator.

In communication applications our interest is in the waveform of the message for which the performance index is more significant, since the attenuator used as reference will bring no new knowledge of the signal waveform. Furthermore, the performance index is simply related to the commonly used noise-to-signal ratio performance. The performance index compares the normalized error of the optimum filter with the normalized error of the optimum attenuator operating on the same input at the same noise level. On the other hand, the noise-to-signal ratio performance compares the noise levels of the optimum filter and the optimum attenuator giving the same error. To establish this second result we need some definitions and derivations concerning noise-to-signal ratio performance. In this we follow the work of Hause.[5]

a. Generalized Noise-to-Signal Ratio Performance

For analytical convenience in the application to the performance of linear filters we shall use the noise-to-signal power ratio instead of the signal-to-noise power ratio, which is commonly found. (Although we call the message the desired output of the filter we shall conserve here the common designation of signal-to-noise ratio. Therefore, the words signal and message are equivalent.) When signal and noise are additive and statistically independent, the noise-to-signal power ratio is commonly defined as

$$\Gamma = \frac{\overline{n^2}}{\overline{m^2}} .$$

Whenever a filter is operating on the input $x(t) = m(t) + n(t)$ and gives an output $y(t)$, then we define a new desired output S such that $S = Cm$, in which C is a constant chosen so that

$$\overline{m(y-Cm)} = 0. \tag{1}$$

The noise is now $(y-Cm)$, and for the noise-to-signal power ratio at the output of the filter we find

$$\Gamma_{out} = \frac{\overline{(y-Cm)^2}}{\overline{C^2m^2}} . \tag{2}$$

This definition, which makes the new signal and the new noise at the output of the filter

linearly independent, reduces to the common definition when the signal and noise are statistically independent. (We use either the noise level k or the noise-to-signal power ratio Γ at the input of the filter for which we have statistical independence, and only the symbol Γ at the output of the filter.) If we make use of the correlation coefficient

$$\rho_{my} = \frac{\overline{my}}{\left(\overline{m^2}\,\overline{y^2}\right)^{1/2}},$$

then we can write Eq. 2 as

$$\Gamma_{out} = \frac{1}{\rho_{my}^2} - 1.$$

Hause[5] has shown that Γ_{out} is minimized by using the optimum mean-square filter followed by an arbitrary amount of gain. For the optimum mean-square filter we can use the known result that the error is uncorrelated with any operation of the same class on the same input to write $\overline{y(m-y)} = 0$, $\overline{my} = \overline{y^2}$, $\overline{e^2} = \overline{m^2} - \overline{y^2}$, and

$$\Gamma_{out} = \frac{\overline{m^2} - \overline{y^2}}{\overline{y^2}} = \frac{1}{(\overline{m^2}/\overline{e^2}) - 1} = \frac{\overline{e^2}/\overline{m^2}}{1 - (\overline{e^2}/\overline{m^2})}. \tag{3}$$

These expressions relate simply the normalized error to the noise-to-signal ratio at the output of an optimum mean-square filter. To connect this result with the behavior of the optimum attenuator we consider the graph of Fig. 2 which gives the normalized error versus the noise level k.

From Eq. 3 we see that for the optimum mean-square filter of some class the noise-to-signal ratio at the output is given by $\frac{A}{B}$. One such filter is the optimum attenuator, and it is clear that for the attenuator the noise-to-signal ratio is k_{att}. Therefore, the
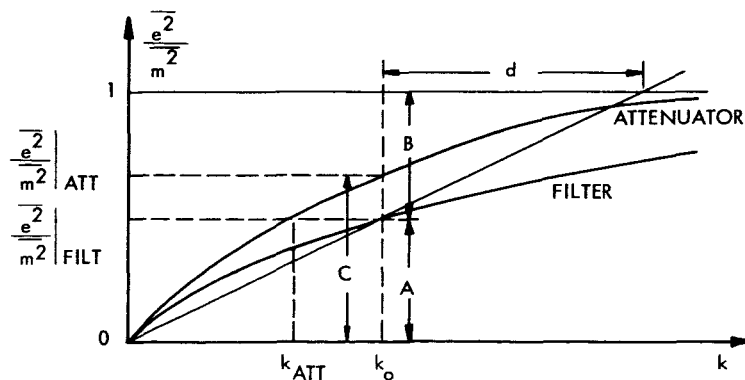
Fig. 2. Normalized error versus noise-to-signal ratio.

output noise-to-signal ratio of an optimum mean-square filter is the input noise level of the optimum attenuator which gives the same normalized error. For a filter operating at the noise level k_o the noise-to-signal ratio performance will be $\dfrac{\Gamma_{out}}{\Gamma_{in}} = \dfrac{k_{att}}{k_o}$, and the performance index will be

$$\eta = \frac{\left.\overline{e^2}/\overline{m^2}\right|_{filt}}{\left.\overline{e^2}/\overline{m^2}\right|_{att}}.$$

A simple geometric construction allows us to find $\dfrac{\Gamma_{out}}{\Gamma_{in}}$ in Fig. 2. Since $\Gamma_{out} = \dfrac{A}{B}$ and $\Gamma_{in} = k_o$, we have $\dfrac{\Gamma_{out}}{\Gamma_{in}} = \dfrac{A}{Bk_o} = \dfrac{1}{d}$, in which d, the distance shown in Fig. 2, is obtained immediately when k_o and the normalized error of the optimum mean-square filter are known. The curves of $\left.\dfrac{\overline{e^2}}{\overline{m^2}}\right|_{filt}$ versus k, which give a constant value to d, are given by

$\left.\dfrac{\overline{e^2}}{\overline{m^2}}\right|_{filt} = \dfrac{k}{d+k}$. For d = 1 we have, as expected, the normalized error of the optimum attenuator.

It appears generally more meaningful to use a reference system operating at the noise level of the filter under study, and we shall therefore emphasize the performance index. We note that by using the same reference system we can define a performance index for non mean-square criteria just as well.

1.4 IMPORTANCE OF THE RELATIVE NOISE LEVEL

As was apparent for the optimum attenuator, the performance of a filter is dependent on the relative noise level. This dependence has not been exploited in filtering and is of interest on both theoretical and practical grounds.

The relative noise level is a parameter that is independent of the signal and noise statistics and is somewhat within our control. If filtering is considered as an alternative to a change of message power or in conjunction with such a change, then the effect of the noise level on the filtering error will be of fundamental importance. In some filtering problems a more appropriate description of the noise might be in terms of stationary statistics with a slowly varying level. If a time-invariant filter is to be used, it is also essential to know how the change of noise level will affect the performance of such a filter. We shall, in several instances in this report, take the noise level to be an explicit parameter and study its effect on the performance of optimum and nonoptimum filters. Our first application is to linear filtering.

II. EFFECT OF NOISE LEVEL ON THE MEAN-SQUARE
PERFORMANCE OF LINEAR FILTERS

In this section we make use of the concepts and parameters discussed in Section I to study the performance of linear systems. In each case the independent parameter will be the noise level. We consider, first, optimum mean-square filters and study the effect of the noise level on the normalized error, the performance index, and the noise-to-signal ratio performance. We then turn to a linear filter that is optimum at some noise level k_o and study its performance at noise levels other than k_o.

2.1 PERFORMANCE OF OPTIMUM LINEAR FILTERS

Here, for simplicity, we shall not take into account the realizability condition. We study, therefore, the irreducible error that corresponds to an arbitrarily large delay in filtering and to the best performance of a linear filter. $H_{opt}(\omega)$, the system function of the optimum linear filter,[6] is

$$H_{opt}(\omega) = \frac{\Phi_{mm}(\omega)}{\Phi_{nn}(\omega) + \Phi_{mm}(\omega)}, \tag{4}$$

in which $\Phi_{mm}(\omega)$ and $\Phi_{nn}(\omega)$ are the power spectra of the message and the noise, respectively. Furthermore, for the normalized mean-square error we have

$$\frac{\overline{e^2}}{\overline{m^2}} = \frac{\int \dfrac{\Phi_{mm}(\omega)\,\Phi_{nn}(\omega)}{\Phi_{nn}(\omega) + \Phi_{mm}(\omega)}\,d\omega}{\int \Phi_{mm}(\omega)\,d\omega}. \tag{5}$$

The integral sign without limits will indicate integration from $-\infty$ to $+\infty$.

Note that if $\Phi_{nn}(\omega) = k\,\Phi_{mm}(\omega)$, we have $H_{opt}(\omega) = \dfrac{1}{1 + k}$, which is the optimum atten-uator, and also $\dfrac{\overline{e^2}}{\overline{m^2}} = \dfrac{k}{1 + k}$.

A rapid but vague estimate of the performance of the optimum linear filter can be obtained by comparing the spectra of message and noise. Limiting cases are clear-cut: If the spectra are identical within a scale factor, we do not get any separation with a linear filter; if they have small overlapping regions, a substantial separation can be expected.

a. Variation of Mean-Square Error with Noise Level

Let $\Phi_{nno}(\omega)$ be such that

$$\int \Phi_{nno}(\omega) = \int \Phi_{mm}(\omega)\,d\omega,$$

7

and let us consider the case for which the shapes of the noise spectrum and the message spectrum are fixed, but the relative noise level can change. Then

$$\Phi_{nn}(\omega) = k\Phi_{nno}(\omega),$$

where k is the relative noise level.

We are interested here in the performance of the linear filter that is optimized at each noise level. The normalized error, expressed in terms of $\Phi_{mm}(\omega)$, $\Phi_{nno}(\omega)$, and k, is

$$\frac{\overline{e^2}}{\overline{m^2}} = \frac{\int \dfrac{k\Phi_{nno}(\omega)\,\Phi_{mm}(\omega)}{k\Phi_{nno}(\omega) + \Phi_{mm}(\omega)}\,d\omega}{\int \Phi_{mm}(\omega)\,d\omega} = \beta(k). \tag{6}$$

We consider this expression to be a function of k only and wish to study some of its properties. For $k = 0$, $\beta(k) = 0$, since the integrand becomes identically zero. If $k \to \infty$, then

$$\beta(k) \to \frac{\int \Phi_{mm}(\omega)\,d\omega}{\int \Phi_{mm}(\omega)\,d\omega} = 1.$$

To obtain this result we require that $\Phi_{mm}(\omega)$ be much less than $k\Phi_{nno}(\omega)$ at all frequencies when k is sufficiently large. This condition is not satisfied at the zeros of $\Phi_{nno}(\omega)$. The whole integrand vanishes, however, at those points and the contribution to the integral of small regions in ω around the zeros of $\Phi_{nno}(\omega)$ can be made arbitrarily small. At all other points the condition $\Phi_{mm}(\omega) \ll k\Phi_{nno}(\omega)$ can be fulfilled for sufficiently large k. (Note that we are excluding the cases in which message and noise are strictly band-limited. For instance, if part of the message bandwidth is free from noise, we get some separation at any noise level.)

$\beta(k)$ is a continuous function of k, since the integrand in the numerator of Eq. 6 is a continuous function of k for all ω.
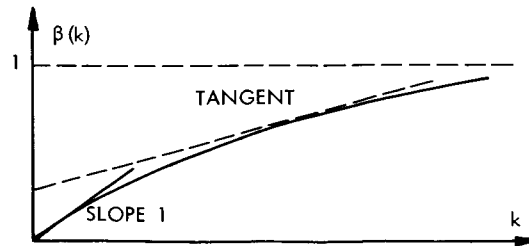
Fig. 3. Normalized error of the optimum linear filter versus noise level.

8

It is shown in Appendix A that the normalized error is a monotonically increasing concave function of the relative noise level k. The rate of increase, however, is always less than the rate of increase of the relative noise level; hence, the slope is ≤ 1. The slope is equal to 1 at the origin, but we know that for no noise at the input the optimum filter becomes

$$H_{opt}(\omega) = \frac{\Phi_{mm}(\omega)}{k\Phi_{nno}(\omega) + \Phi_{mm}(\omega)}\bigg|_{k=0} = 1.$$

For such a device $\beta(k) = k$ and $d\beta/dk = 1$. We note for future use that any tangent to the $\beta(k)$ curve will have a slope ≤ 1 and will intersect the $\beta(k)$ axis between zero and one. $\beta(k)$ can be sketched as shown in Fig. 3.

b. Performance Index

We defined the performance index for an optimum filter as the ratio of its normalized error to the normalized error of the optimum attenuator. If we consider this index as a function of the relative noise level k, we write

$$\frac{\overline{e^2/m^2}\big|_{filt}}{\overline{e^2/m^2}\big|_{att}} = \frac{\beta(k)}{\frac{k}{1+k}} = \frac{1+k}{k}\beta(k) = \eta(k)$$

$$\eta(k) = \frac{(1+k)\int \dfrac{\Phi_{nno}(\omega)\,\Phi_{mm}(\omega)}{k\Phi_{nno}(\omega) + \Phi_{mm}(\omega)}\,d\omega}{\int \Phi_{mm}(\omega)\,d\omega}.$$

For k = 0 we write

$$\eta(0) = \frac{\int \Phi_{nno}(\omega)\,d\omega}{\int \Phi_{mm}(\omega)\,d\omega} = 1.$$

For k $\to \infty$ under conditions mentioned before we can write

$$\eta(k) \underset{k\to\infty}{\to} \frac{\int \Phi_{mm}(\omega)\,d\omega}{\int \Phi_{mm}(\omega)\,d\omega} = 1.$$

Thus at zero and very large noise levels the optimum linear filter cannot perform any better than the optimum attenuator.

We wish to show that $\eta(k) \leq 1$ for all k, and this requires

9

$$\int \frac{(1+k)\,\Phi_{nno}(\omega)\,\Phi_{mm}(\omega)}{k\Phi_{nno}(\omega) + \Phi_{mm}(\omega)}\,d\omega \leq \int \Phi_{mm}(\omega)\,d\omega,$$

which can be written

$$\int \frac{\dfrac{\Phi_{mm}(\omega)}{\Phi_{nno}(\omega)}\left[\Phi_{nno}(\omega) - \Phi_{mm}(\omega)\right]}{k + \dfrac{\Phi_{mm}(\omega)}{\Phi_{nno}(\omega)}}\,d\omega \leq 0. \tag{7}$$

Let

$$\frac{\dfrac{\Phi_{mm}(\omega)}{\Phi_{nno}(\omega)}}{k + \dfrac{\Phi_{mm}(\omega)}{\Phi_{nno}(\omega)}} = a.$$

Inequality (7) is then written

$$\int a\left[\Phi_{nno}(\omega) - \Phi_{mm}(\omega)\right]\,d\omega \leq 0. \tag{8}$$

If a is a constant, then inequality (8) reduces to an equality from the definition of $\Phi_{nno}(\omega)$. If the integrand of (8) is positive, that is, $\Phi_{nno}(\omega) \geq \Phi_{mm}(\omega)$, then $a \leq \dfrac{k}{1+k}$, and if the integrand is negative, we have $a \geq \dfrac{k}{1+k}$. It is therefore clear that inequality (8) is satisfied and $\eta(k) \leq 1$ for all k.

By a similar reasoning it is possible to show that the slope of $\eta(k)$ at the origin is always negative. This slope at k = 0 is

$$\left.\frac{d\eta(k)}{dk}\right|_{k=0} = \frac{\displaystyle\int \frac{\Phi_{nno}(\omega)}{\Phi_{mm}(\omega)}\left[\Phi_{mm}(\omega) - \Phi_{nno}(\omega)\right]\,d\omega}{\displaystyle\int \Phi_{mm}(\omega)\,d\omega}.$$

This initial slope gives an indication of the performance index at low noise levels.

Since we have found that $\eta(k)$ is a continuous function of k, which is always ≤ 1 for all k and goes to one for k = 0 and k $\to \infty$, $\eta(k)$ will go through one or several minima for $0 \leq k < \infty$.

c. Output Noise-to-Signal Ratio

We established in Section I that the noise-to-signal ratio at the output of an optimum

mean-square filter of some class is given by

$$\Gamma = \frac{\overline{e^2/m^2}}{1 - \overline{e^2/m^2}} \; .$$

We apply this result to the optimum linear filter and study the dependence of Γ on k, the noise level. If $\beta(k)$ denotes $\overline{e^2}/\overline{m^2}$ as before, we can write

$$\Gamma(k) = \frac{\beta(k)}{1 - \beta(k)} = \frac{1}{1 - \beta(k)} - 1.$$

From the properties of $\beta(k)$ we immediately have the facts that $\Gamma(k)$ will be a monotonically increasing function of k and that $\Gamma(0) = 0$, $\Gamma(k) \to \infty$ as $k \to \infty$. Furthermore, we have

$$\frac{d\Gamma(k)}{dk}\bigg|_{k=0} = \frac{\beta'(0)}{[1-\beta(0)]^2} = 1 \tag{9a}$$

and

$$\frac{d\Gamma(k)}{dk}\bigg|_{k\to\infty} \to \frac{\int \Phi_{mm}(\omega)\, d\omega}{\int \dfrac{\Phi_{mm}^2(\omega)}{\Phi_{nno}(\omega)}\, d\omega} \leq 1. \tag{9b}$$

Equations 9 can be established without difficulty. It is slightly more difficult to show that $\Gamma(k)$ is a concave function of k, and this is proved in Appendix B.

d. Noise-to-Signal Ratio Performance

This common measure of performance compares the noise-to-signal ratios at the input and the output of the filter. We have

$$\frac{\Gamma_{out}}{\Gamma_{in}} = \frac{\beta(k)}{k[1-\beta(k)]} = \frac{\displaystyle\int \frac{\Phi_{nno}(\omega)\, \Phi_{mm}(\omega)}{k\Phi_{nno}(\omega) + \Phi_{mm}(\omega)}\, d\omega}{\displaystyle\int \frac{\Phi_{mm}^2(\omega)}{k\Phi_{nno}(\omega) + \Phi_{mm}(\omega)}\, d\omega} \; .$$

It can be shown easily that $\dfrac{\Gamma_{in}}{\Gamma_{out}}\bigg|_{k=0} = 1$, and that for $k \to \infty$ we have

11

$$\frac{\Gamma_{out}}{\Gamma_{in}} \to \frac{\int \Phi_{mm}(\omega)\, d\omega}{\int \dfrac{\Phi^2_{mm}(\omega)}{\Phi_{nno}(\omega)}\, d\omega} \leq 1. \tag{10}$$

Expression (10) is clearly equal to the slope of Γ_{out} as $k \to \infty$. The slope of $\dfrac{\Gamma_{out}}{\Gamma_{in}}$ at the
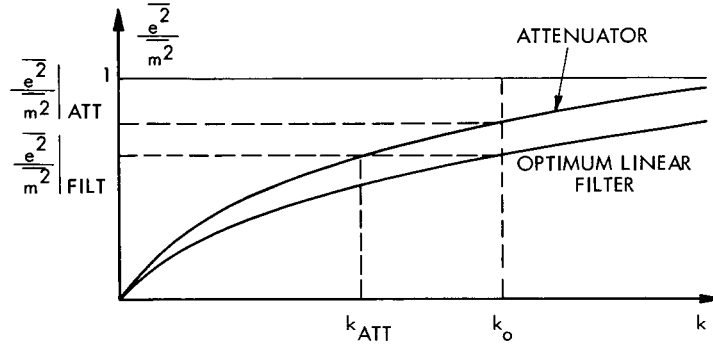
Fig. 4. Performance of linear filters for large noise-to-signal ratios.

origin, which gives an indication of the performance at low noise levels, is

$$\frac{d}{dk} \left. \frac{\Gamma_{out}}{\Gamma_{in}} \right|_{k=0} = \frac{\int \dfrac{\Phi^2_{mm}(\omega)}{\Phi_{nno}(\omega)}\, d\omega \ \int \dfrac{\Phi^2_{nno}(\omega)}{\Phi_{mm}(\omega)}\, d\omega}{\left[\int \Phi_{mm}(\omega)\, d\omega \right]^2} - 1.$$

It is of interest to note an apparent discrepancy between the noise-to-signal ratio performance and the performance index. We found earlier that $\eta(k) \to 1$ for $k \to \infty$ and, therefore, that the optimum linear filter does not perform better than the optimum attenuator as $k \to \infty$. We find here that $\dfrac{\Gamma_{out}}{\Gamma_{in}}$ can be substantially less than 1 for $k \to \infty$.

Since the optimum attenuator leads to $\dfrac{\Gamma_{out}}{\Gamma_{in}} = 1$ for all k, this seems to contradict the previous result. This difference is explained by considering the graph of Fig. 4.

Since $\eta(k_o) = \dfrac{\overline{e^2/m^2}\big|_{filt}}{\overline{e^2/m^2}\big|_{att}}$ and $\dfrac{\Gamma_{out}}{\Gamma_{in}}(k_o) = \dfrac{k_{att}}{k_o}$, we see that, as $k_o \to \infty$, $\eta(k_o) \to 1$ but $\dfrac{k_{att}}{k_o}$

can be less than 1. That is, a substantial change of input noise level is needed before the optimum attenuator gives the incrementally lower normalized error of the optimum

linear filter, but this is clearly an illusory improvement.

2.2 PERFORMANCE OF NONOPTIMUM LINEAR FILTERS

We show in this section that both the normalized error and the output noise-to-signal ratio of an arbitrary linear filter are linear functions of the noise level k. If the linear filter is optimum at noise level k_o, then the tangents to $\beta_{opt}(k)$ and $\Gamma_{opt}(k)$ at point $k = k_o$ will give its behavior for $k \neq k_o$.

a. Normalized Error of a Linear Filter

Let $h(t)$ be the impulse response of an arbitrary linear system, $H(\omega)$ the system function, and $x(t)$ and $y(t)$ the input and the output, respectively. We have $x(t) = m(t) + n(t)$, as before, and we are interested in the performance of this system as a filter for the extraction of $m(t)$.

$$y(t) = \int h(t)\, x(t-\tau)\, d\tau = m_1(t) + n_1(t)$$

$$m_1(t) \overset{\Delta}{=} \int h(t)\, m(t-\tau)\, d\tau$$

$$n_1(t) \overset{\Delta}{=} \int h(\tau)\, n(t-\tau)\, d\tau .$$

If the message alone is applied to the input, the mean-square error becomes

$$\overline{e_m^2} = \overline{(m_1-m)^2}.$$

If both message and noise are applied to the input,

$$\overline{e_{m+n}^2} = \overline{(y-m)^2} = \overline{(m_1+n_1-m)^2} = \overline{(m_1-m)^2} + \overline{n_1^2} + 2\overline{n_1(m_1-m)},$$

but

$$\overline{n_1 m_1} = \int\int h(\tau_1)\, h(\tau_2)\, \overline{m(t-\tau_1)\, n(t-\tau_2)}\, d\tau_1\, d\tau_2$$

and

$$\overline{m(t-\tau_1)\, n(t-\tau_2)} = 0,$$

since message and noise are uncorrelated and have zero mean. Hence

$$\overline{n_1 m_1} = 0,$$

and, similarly,

13

$$\overline{n_1 m} = 0$$

and

$$\overline{e_{m+n}^2} = \overline{(m_1 - m)^2} + \overline{n_1^2} = \overline{e_m^2} + \overline{n_1^2}.$$

Here, $\overline{n_1^2}$ is the mean-square value of the output of the system when noise alone is applied to it.

If we define $n_o(t)$ by

$$n(t) = \sqrt{k}\, n_o(t)$$

and

$$\overline{n_o^2} = \overline{m^2},$$

then the mean-square value of the output of the system when $n_o(t)$ is applied to it is

$$\overline{\left[\int h(\tau)\, n_o(t-\tau)\, d\tau\right]^2} = \int \Phi_{nno}(\omega)\, |H(\omega)|^2\, d\omega = C.$$

Here, C is a constant,

$$\overline{n_1^2} = \overline{\left[\int h(\tau)\, n(t-\tau)\, d\tau\right]^2} = Ck \tag{11}$$

and

$$\overline{e_{m+n}^2} = \overline{e_m^2} + Ck.$$

The mean-square error varies linearly with the relative noise level k. The normalized error is

$$\frac{\overline{e_{m+n}^2}}{\overline{m^2}} = \frac{\overline{e_m^2}}{\overline{m^2}} + \frac{C}{\overline{m^2}}\, k.$$

Let us consider the case for which the linear system is the optimum filter at a noise level k_o. The system function is then

$$H(\omega) = \frac{\Phi_{mm}(\omega)}{k_o \Phi_{nno}(\omega) + \Phi_{mm}(\omega)},$$

and the slope of the normalized error versus k becomes

$$\frac{C}{\overline{m^2}} = \frac{\int \Phi_{nno}(\omega) \frac{[\Phi_{mm}(\omega)]^2}{[k_o \Phi_{nno}(\omega) + \Phi_{mm}(\omega)]^2} \, d\omega}{\int \Phi_{mm}(\omega) \, d\omega} .$$ (12)

If we compare (12) with the slope of the optimum normalized error curve at point k_o, it is easy to verify the fact that

$$\frac{C}{\overline{m^2}} = \frac{d\beta(k)}{dk} \bigg|_{k=k_o} .$$

Thus, the normalized error of the filter designed to be optimum at noise level k_o is obtained simply by drawing the tangent to the curve of normalized error $\beta_{opt}(k)$ at point k_o. The fact that $\beta(k)$ is a concave function of k (Appendix B) ensures that the normalized error for this filter will be larger than optimum for all $k \neq k_o$. Furthermore, since all tangents to $\beta(k)$ intersect the $\beta(k)$ axis between zero and one, we see that for any linear filter the distortion in the absence of noise is always larger than zero and it is equal to zero only if $k_o = 0$. Therefore we cannot separate the signal from the noise unless some distortion of the message in the absence of noise is accepted.

b. Output Noise-to-Signal Ratio of a Linear Filter

We have seen that the output noise-to-signal ratio for a filter, whether or not it is optimum in the mean-square sense, is given by $\Gamma_{out} = \left(1/\rho_{my}^2\right) - 1$. For a linear filter we write, as before, $y = m_1 + n_1$ and we have, furthermore, $\overline{my} = \overline{mm_1}$ and $\overline{y^2} = \overline{m_1^2} + \overline{n_1^2}$. Therefore we have

$$\rho_{my}^2 = \frac{\overline{mm_1}^2}{\overline{m^2}\left(\overline{m_1^2} + \overline{n_1^2}\right)} .$$

We showed (Eq. 11) that $\overline{n_1^2} = Ck$, in which C is a constant related to the system function and the noise spectrum. Thus the output noise-to-signal ratio for a linear filter takes the form

$$\Gamma_{out} = \frac{\overline{m^2}\left(\overline{m_1^2} + Ck\right)}{\overline{mm_1}^2} - 1 .$$

Therefore, Γ_{out} is a linear function of k, and we note that the output noise-to-signal ratio for $k = 0$ is $\Gamma_{out}(0) = \left(\overline{m^2}\,\overline{m_1^2}/\overline{mm_1}^2\right) - 1$, which depends only on the filter and the

message spectrum. It can be verified easily that if the filter is optimum at noise level k_o, then its output noise-to-signal ratio versus k for $k \neq k_o$ is the tangent to $\Gamma_{opt}(k)$ at $k = k_o$. Since $\Gamma_{opt}(k)$ is a concave function, such a nonoptimum filter will give an output noise-to-signal ratio larger than optimum for $k \neq k_o$.

2.3 EXAMPLE

To illustrate the various points discussed, we consider

$$\Phi_{nn}(\omega) = \frac{D^2 k}{1 + \omega^2} \qquad \Phi_{mm}(\omega) = \frac{E^2 \omega^2}{(1+\omega^2)(4+\omega^2)},$$

where E^2 is chosen so that $\int \Phi_{mm}(\omega) \, d\omega = 2\pi$, and D^2 is found by writing $\int \Phi_{nn}(\omega) \, d\omega = 2\pi$ for $k = 1$; this gives $E^2 = 6$ and $D^2 = 2$.
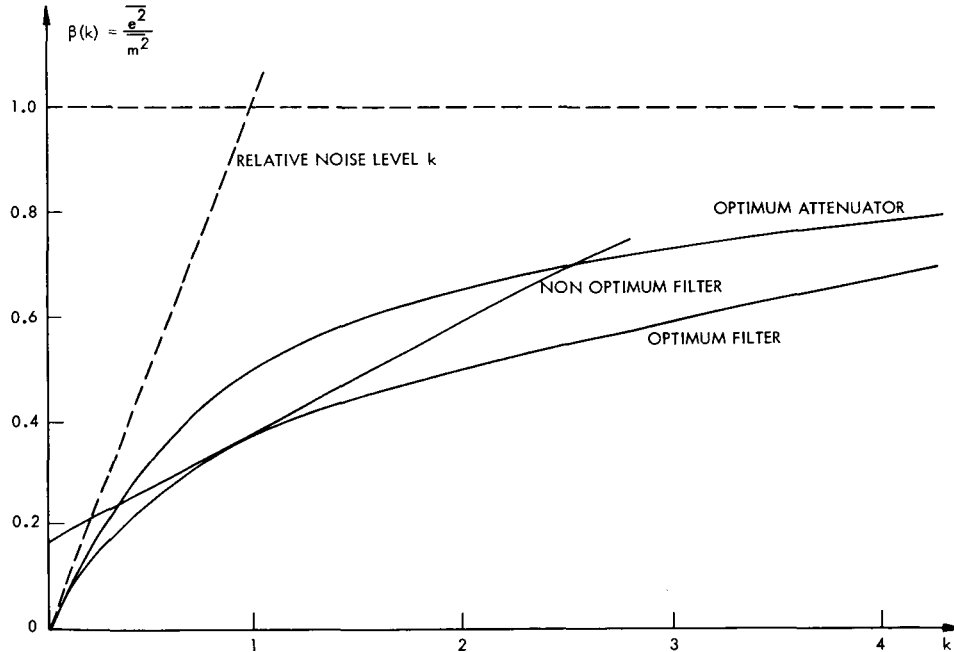
Fig. 5. Normalized error of optimum and nonoptimum linear filters versus noise level.

Of the various quantities defined, only the normalized error requires computation.

$$\beta_{opt}(k) = \frac{\int \dfrac{k\Phi_{nno}(\omega)\, \Phi_{mm}(\omega)}{k\Phi_{nno}(\omega) + \Phi_{mm}(\omega)} \, d\omega}{\int \Phi_{mm}(\omega) \, d\omega}.$$

16
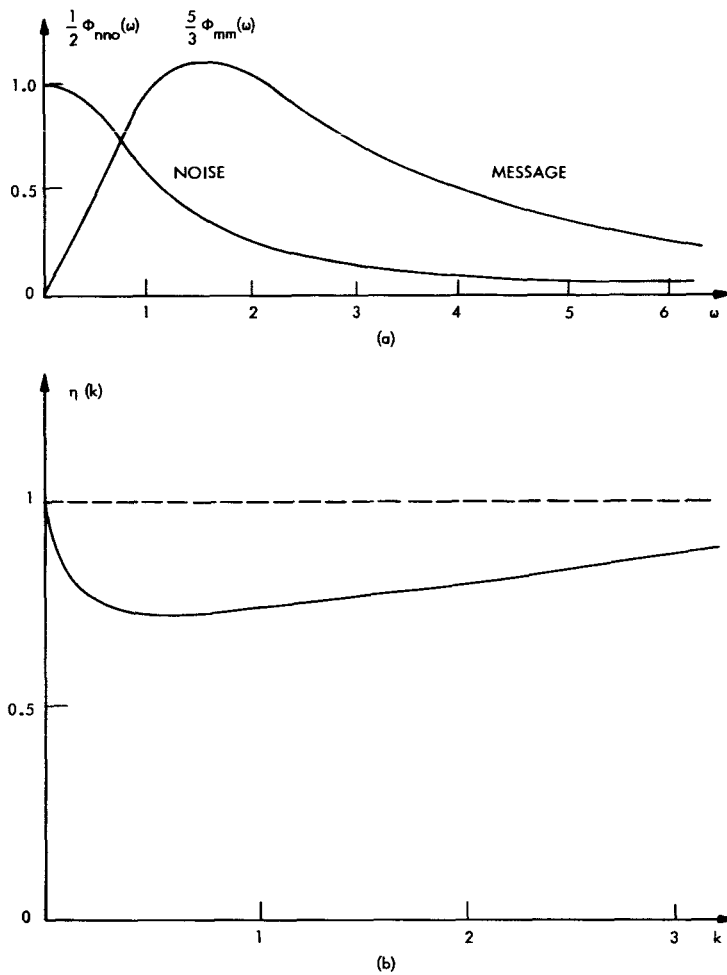
Fig. 6. (a) Message and noise power spectra. (b) Performance index of the optimum linear filter.

The computation of this definite integral is carried out by the use of the tables of Newton, Gould, and Kaiser.[7] Our results are summarized.

Optimum filter

Normalized error

$$\beta(k) = \frac{3k}{2\sqrt{k(k+3)} + k + 3}$$

Performance index

$$\eta(k) = \frac{3(1+k)}{2\sqrt{k(k+3)} + k + 3}$$

Output noise-to-signal ratio

$$\Gamma(k) = \frac{3k}{2\sqrt{k(k+3)} + 3 - 2k}$$

Noise-to-signal ratio performance

$$\frac{\Gamma_{out}}{\Gamma_{in}} = \frac{3}{2\sqrt{k(k+3)} + 3 - 2k}$$
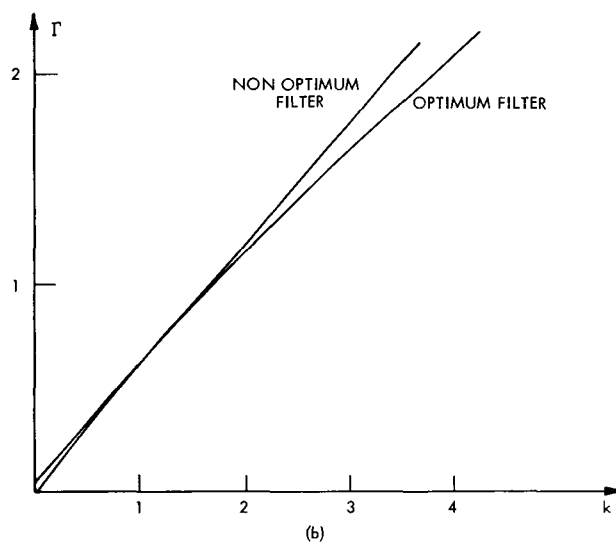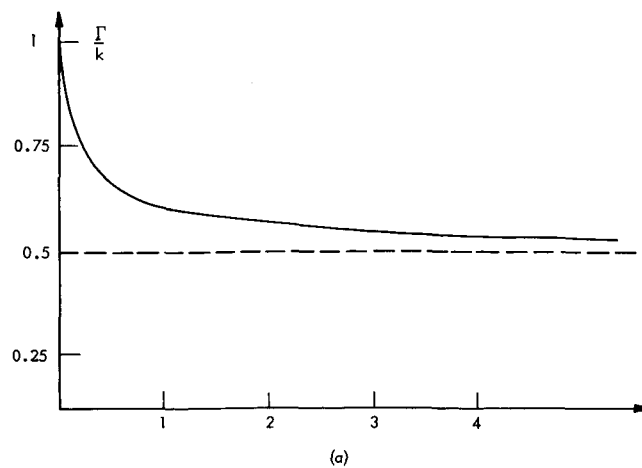
Fig. 7. (a) Noise-to-signal ratio performance of the optimum filter.
(b) Output noise-to-signal ratios.

Nonoptimum filter (optimum for k = 1)

Normalized error $\qquad \beta(k) = \dfrac{3}{128}\,(9k+7)$

Output noise-to-signal ratio $\qquad \Gamma(k) = 0.54k + 0.06$

These results are represented graphically in Figs. 5, 6, and 7.

III. OPTIMUM NONLINEAR NO-MEMORY FILTER

In this section we consider the problem of finding the characteristic of the optimum nonlinear no-memory filter, and examine some properties of such filters. Most of the results presented here are not new and are generally mentioned in connection with the theory of estimation. (See, for instance, Blackwell and Girshick.[8]) We present them here for the sake of completeness and clarity, and we shall draw on them later in our discussion of the performance of nonlinear no-memory filters.

3.1 EXPRESSION FOR THE OPTIMUM NONLINEAR NO-MEMORY FILTER

The filtering operation considered and the notation used are shown in Fig. 8. We call $W(e)$ the error-weighting function, and we wish to find $g(x)$, which minimizes the
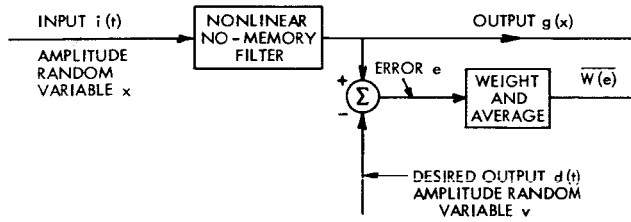
Fig. 8. Illustrating notation.

average weighted error, $E[W(e)] \triangleq \overline{W(e)}$. We write

$$\overline{W(e)} = \iint W[v-g(x)]\, p_{d,i}(v,x)\, dv\, dx = \int \left[\int W[v-g(x)]\, p_{d/i}(v/x)\, dv \right] p_i(x)\, dx. \tag{13}$$

Since the filter characteristic to be found, $g(x)$, is only a function of the input amplitude, x, it is evident that if we choose $g(x)$ such that for each x the integral in the bracket of Eq. 13 is minimum, then $\overline{W(e)}$ will be minimum. Hence, we wish to find $g(x)$ such that it minimizes

$$\overline{W_x(e)} \triangleq \int W[v-g(x)]\, p_{d/i}(v/x)\, dv,$$

where $\overline{W_x(e)}$ is the average weighted error, given that the input is x. The minimizing $g(x)$ is found by equating to zero the variation of $\overline{W_x(e)}$ with respect to $g(x)$.

$$\delta \overline{W_x(e)} = \int \frac{\partial\{W[v-g(x)]\}}{\partial g}\, \delta g(x)\, p_{d/i}(v/x)\, dx = 0,$$

and this should hold for any $\delta g(x)$.

If we let $\partial W(e)/\partial g = -\, dW(e)/de = -f(e)$, then $g(x)$ is determined by solving

$$\int f[v-g(x)]\, p_{d/i}(v/x)\, dv = 0. \tag{14}$$

Once the conditional probability density, $p_{d/i}(v/x)$, is known for all x, Eq. 14 allows us to find formally the filter $g(x)$ corresponding to an arbitrary criterion $W(e)$. In some

cases a more explicit relation for $g(x)$ can be obtained. We consider five criteria for illustration and later use.

a. Mean Absolute Value of the Error Criterion

We take

$$W(e) = |e|$$

and, hence,

$$f(e) = \frac{dW(e)}{de} = \text{sgn } e,$$

and Eq. 14 becomes

$$\int \text{sgn}\,[v-g(x)]\,p_{d/i}(v/x)\,dv = 0$$

or

$$\int_{-\infty}^{g(x)} p_{d/i}(v/x)\,dv = \int_{g(x)}^{\infty} p_{d/i}(v/x)\,dv,$$

which is clearly equivalent to

$$\int_{-\infty}^{g(x)} p_{d/i}(v/x) = 1/2.$$

Hence, $g(x)$ is the median of the conditional probability density of the desired output, given the input. This criterion was considered recently in detail by Bluestein and Schwarz[2] in connection with the problem of signal quantization. It has the interesting property that if the desired output $d(t)$ is quantized, the optimum output is quantized at the same levels.

b. Mean-Square Error Criterion

Since

$$W(e) = e^2 \qquad \text{and} \qquad f(e) = 2e,$$

we write Eq. 14

$$\int [v-g(x)]\,p_{d/i}(v/x)\,dv = 0$$

$$g(x) = \int v\,p_{d/i}(v/x)\,dv.$$

We have the classical result that the optimum filter is now given by the mean of the conditional probability density.

c. Mean Fourth Power of the Error Criterion

$$W(e) = e^4$$

and now

$$f(e) = 4e^3,$$

which for Eq. 14 leads to

$$\int [v-g(x)]^3 \, p_{d/i}(v/x) \, dv = 0$$

or

$$\int [v^3 - 3v^2 g(x) + 3vg^2(x) - g^3(x)] \, p_{d/i}(v/x) \, dv = 0.$$

Let us define a symbol for the nth moment of the conditional probability density as

$$\int v^n \, p_{d/i}(v/x) \, dv = \overline{d_x^n}.$$

Then we have

$$g^3(x) - 3\overline{d_x}\,g^2(x) + 3\overline{d_x^2}\,g(x) - \overline{d_x^3} = 0. \tag{15}$$

The optimum filter characteristic is found as a root of this cubic equation (15) in $g(x)$ which involves only the moments of the conditional probability density. It is a simple matter to show that Eq. 15 has only one real root. For this, consider $\ell[g(x)]$ the derivative of the left-hand side of Eq. 15 with respect to $g(x)$.

$$\ell[g(x)] = 3g^2(x) - 6\overline{d_x}\,g(x) + 3\overline{d_x^2}.$$

Here, $\ell[g(x)]$ has no real roots, since the variance of the conditional probability density $\sigma_{dx}^2 = \overline{d_x^2} - \overline{d_x}^2$ is always positive. This, in turn, implies that Eq. 15 has only one real root.

Note that if we take $g(x) = \overline{d_x}$, then Eq. 15 can be written as $\overline{(v-\overline{d_x})^3} = 0$ and will be satisfied if the third central moment of $p_{d/i}(v/x)$ is zero; if such is the case, the conditional mean will give the optimum mean fourth power of the error filter.

d. Error Criterion Characterized by $W(e) = \begin{cases} 0 & |e| < A \\ 1 & |e| \geq A \end{cases}$

This criterion does not penalize error smaller than $|A|$ and counts with equal weight all errors larger than $|A|$. The average weighted error $\overline{W(e)}$ is, therefore, the probability that the error is larger than $|A|$, and the optimum filter minimizes this probability.

For this criterion we have

$$f(e) = \frac{dW(e)}{de} = u(e-A) - u(e+A),$$

in which u( ) denotes a unit impulse. Equation 15 becomes

$$\int \left\{ u[v-g(x)-A] - u[v-g(x)+A] \right\} p_{d/i}(v/x) \, dv = 0$$

or

$$p_{d/i}[A+g(x)/x] = p_{d/i}[g(x)-A/x].$$

If for a given x we use p(v) instead of $p_{d/i}(v/x)$, then the constant C = g(x) has to be such that

$$p(C-A) = p(C+A). \tag{16}$$

This relation may have several solutions. The physical situation is clear: We wish to encompass the largest total probability between the points C – A and C + A, and relation (16) corresponds to stationary values of the probability in an interval of length 2A when the conditional probability density is continuous. Note that if the conditional probability density is even, continuous, and has its only peak at the mean, then the conditional mean gives the optimum filter.

e. Error Criterion Characterized by W(e) = cosh e–1

We consider this criterion for illustration purposes and obtain

$$f(e) = \sinh e.$$

For Eq. 14 we now have

$$\int \sinh [v-g(x)] \, p_{d/i}(v/x) \, dv = 0. \tag{17}$$

Expanding the hyperbolic sine, we have

$$\cosh [g(x)] \int \sinh v \, p_{d/i}(v/x) \, dv - \sinh [g(x)] \int \cosh v \, p_{d/i}(v/x) \, dv = 0,$$

which yields

$$\tanh [g(x)] = \frac{\int \sinh v \, p_{d/i}(v/x) \, dv}{\int \cosh v \, p_{d/i}(v/x) \, dv}.$$

Let us designate $P_x(jt)$ the characteristic function of the conditional probability density

$$P_x(jt) = \int e^{jtv} \, p_{d/i}(v/x) \, dv.$$

If $P_x(jt)$ is analytic in the complex s plane ($s = \sigma + jt$) out to points $s = \pm 1$, then we can write

$$g(x) = \tanh^{-1}\left[\frac{P_x(-1) - P_x(1)}{P_x(-1) + P_x(1)}\right].$$ (18)

We can note here too that if $p_{d/i}(v/x)$ is even around the conditional mean $v_o = \overline{d_x}$, then we satisfy Eq. 17 by taking $g(x) = \overline{d_x}$, and this is then the unique solution.

For easy comparison we give in Fig. 9 a graph of the various error-weighting functions considered. From a practical viewpoint, the absolute value of the error criterion and criterion d are of greatest interest after the mean-square criterion.

Thus we have illustrated how the optimum filter for non mean-square criteria, as well as for the mean-square criterion, can be found for nonlinear no-memory filters. This optimum filter is sometimes expressed only in terms of moments of the conditional probability density.

In any case, the only information required is the conditional probability density for the desired output, given the input. The expressions obtained are, in fact, valid for the case of filters with memory, but the conditional probability density then depends on the past of the input. Only for filters without memory, however, can the conditional probability density be easily obtained.

3.2 UNIQUENESS OF THE OPTIMUM FILTER

A sufficient condition to ensure uniqueness of the solution of Eq. 14, and hence a unique optimum filter, is to require that $W(e)$, the error-weighting function, be convex. If $W(e)$ is convex, then $f(e) = \dfrac{dW(e)}{de}$ is a monotonically increasing function of e.

Consider Fig. 10, which illustrates this case. If $f[v-g(x)]$ is a monotonically
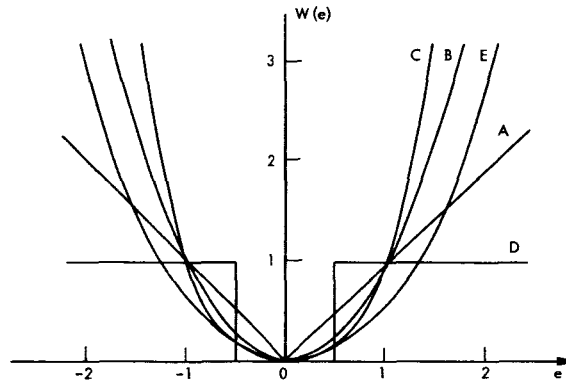
Fig. 9. Error-weighting functions considered.

23
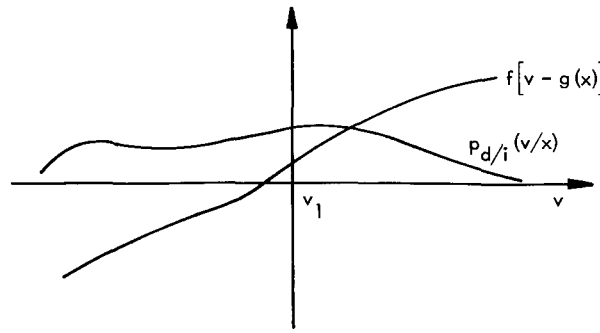
Fig. 10. Illustration of the conditions for uniqueness of the optimum filter.

increasing function of v, then both

$$\int_{v_1}^{\infty} f[v-g(x)] \, p_{d/i}(v/x) \, dv$$

and

$$\int_{-\infty}^{v_1} f[v-g(x)] \, p_{d/i}(v/x) \, dv$$

are monotonically decreasing functions of v_1 and, therefore, of $-g(x)$. Hence,

$$\int_{-\infty}^{+\infty} f[v-g(x)] \, p_{d/i}(v/x) \, dv$$

is a monotonically increasing function of $g(x)$, and Eq. 14 has only one solution. This result holds without any restriction on the conditional probability density, $p_{d/i}(v/x)$.

3.3 EQUIVALENCE OF MEAN-SQUARE AND NON MEAN-SQUARE FILTERING

For a wide class of error criteria and some types of conditional probability densities, Sherman[9] has established an important property on the equivalence of mean-square filtering with non mean-square filtering.

In our notation his result takes the following form: If the error-weighting function is of the form

$$0 \leq W(e) = W(-e)$$

and

$$0 \leq e_1 \leq e_2 \rightarrow W(e_1) \leq W(e_2)$$

and if $p_{d/i}(v/x)$ is even about $v = v_0$, does not contain any impulses, and is monotonically increasing for $v < v_0$, then the minimum average weighted error is obtained by taking $v^* = v_0$ as an estimate.

This result leads to the conclusion that whenever the input and the desired output are

24

Gaussian, and, therefore, $p_{d/i}(v/x)$ is Gaussian, then the best filter is the same for any member of the class of criteria stated. Since the best filter is linear for the mean-square error criterion, it is linear for any of the other criteria. Note, however, that Sherman's result applies to conditional probability densities that are not Gaussian.

Without calling upon the results of Sherman, it is simple to see that the five criteria considered previously lead to the same filter when $p_{d/i}(v/x)$ is Gaussian. It is possible, in fact, to relax Sherman's requirements for some error criteria.

Consider again Eq. 14. In order to determine the optimum filter, we have to find a constant $g(x)$ such that (14) is satisfied for the specific value of the input considered. Let $f(e) = \dfrac{dW(e)}{de}$ be an odd function of e, non-negative for $e \geqslant 0$. Then if $p_{d/i}(v/x)$ is even about $v = v_o$, $g(x) = v_o$ is clearly a solution of Eq. 14. If we have a unique solution, then the conditional mean will correspond to the minimum of the average weighted error for all criteria satisfying the conditions given above. To ensure a unique solution to Eq. 14, further conditions are needed, either on the error-weighting function $W(e)$ or on the conditional probability density, $p_{d/i}(v/x)$.
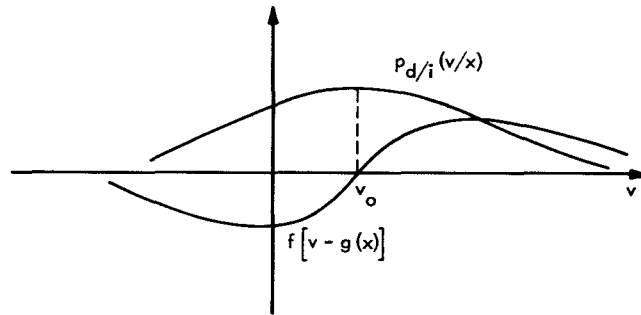
Fig. 11. Conditional probability density for which the non mean-square filter is the same as the mean-square filter.

One such sufficient condition on the conditional probability density is to require that $p_{d/i}(v/x)$ be monotonically increasing for $v < v_o$, and this is Sherman's result. That this is sufficient can be seen by considering the graph for such a case given in Fig. 11. It is seen that Eq. 14 will not be satisfied by any $g(x) \neq v_o$.

We have stated that a convex error-weighting function $W(e)$ will ensure a unique solution of Eq. 14 for any conditional probability density. We have the following result: If the conditional probability density $p_{d/i}(v/x)$ is even about $v = v_o$, then $g(x) = v_o$ is the optimum estimate for all error-weighting functions that are even and convex.

Some of the error criteria considered earlier, which do not satisfy the convexity requirement, will lead to nonunique optimum filters for some conditional probability densities.

25

IV. NO SEPARATION BY A NONLINEAR NO-MEMORY FILTER

We consider in this section the classes of message and noise for which the optimum nonlinear no-memory filter reduces to an attenuator. From our point of view we say that in those cases the message and the noise cannot be separated by a nonlinear no-memory filter. In this discussion, the major emphasis will be given to filtering in the mean-square sense, but the results on the equivalence of mean-square and non mean-square filtering given in Section III will allow some extension to non mean-square filtering. Since the input i(t) is the sum of the message m(t) and the noise n(t), which are statistically independent, and since the desired output d(t) is the message m(t), we have

$$p_{d/i}(v/x) = p_{m/m+n}(v/x)$$

and

$$p_{m/m+n}(v/x) = \frac{p_{m+n/m}(x/v)\, p_m(v)}{p_{m+n}(x)}$$

$$p_{m/m+n}(v/x) = \frac{p_n(x-v)\, p_m(v)}{p_{m+n}(x)}. \tag{19}$$

Equation 19 is the expression for the conditional probability density that we shall use in our discussion.

4.1 THE OPTIMUM NONLINEAR NO-MEMORY FILTER IN THE MEAN-SQUARE SENSE REDUCED TO AN ATTENUATOR

The optimum mean-square filter is given by the conditional mean

$$g(x) = \int v\, p_{m/m+n}(v/x)\, dv \tag{20}$$

which, with the use of Eq. 19, can be written

$$g(x) = \frac{\int v\, p_n(x-v)\, p_m(v)\, dv}{p_{m+n}(x)}. \tag{21}$$

Let

$$r(x) = \int v\, p_n(x-v)\, p_m(v)\, dv$$

$$q(x) = \int p_n(x-v)\, p_m(v)\, dv.$$

Equation 21 can be written

$$g(x) = \frac{r(x)}{q(x)}.$$

(22)

We shall now express $r(x)$ and $g(x)$ in terms of the characteristic functions $P_n(t)$ and $P_m(t)$. We use the definitions

$$p(v) = \frac{1}{2\pi} \int P(t) \, e^{-jtv} \, dt$$

$$P(t) = \int p(v) \, e^{jtv} \, dv.$$

Writing $r(x)$ in terms of $P_m(t)$ and $P_n(t)$, we obtain

$$r(x) = \iint v \, p_m(v) \, \frac{P_n(t)}{2\pi} \, e^{-jt(x-v)} \, dt dv = \frac{1}{2\pi} \int P_n(t) \, e^{-jtx} \, dt \int v \, p_m(v) \, e^{jtv} \, dv$$

$$= \frac{1}{2\pi} \int P_n(t) \, \frac{dP_m(t)}{d(jt)} \, e^{-jtx} \, dt$$

and similarly for $q(x)$ we have

$$q(x) = \frac{1}{2\pi} \int P_n(t) \, P_m(t) \, e^{-jtx} \, dt.$$

Hence,

$$g(x) = \frac{\int P_n(t) \, \dfrac{dP_m(t)}{d(jt)} \, e^{-jtx} \, dt}{\int P_n(t) \, P_m(t) \, e^{-jtx} \, dt}.$$

(23)

Equation 23 is the expression that we need for our discussion. If the optimum filter is $g(x) = ax + b$, an attenuator, then Eq. 22 becomes

$$(ax+b) \, q(x) = r(x).$$

(24)

By the use of Eq. 23 we can equate the Fourier transforms of Eq. 24 and, in terms of characteristic functions, we write

$$a \, \frac{d}{d(jt)} \, [P_m(t) \, P_n(t)] + b[P_m(t) \, P_n(t)] = P_n(t) \, \frac{dP_m(t)}{d(jt)}$$

or

$$(1-a) \, P_n(t) \, \frac{dP_m(t)}{d(jt)} - a \, P_m(t) \, \frac{dP_n(t)}{d(jt)} - b \, P_n(t) \, P_m(t) = 0.$$

(25)

27

This linear differential equation (25) relating the characteristic functions of the message and the noise was obtained by Balakrishnan[10] who used a different derivation. The solution of (25) is

$$P_m(t) = [P_n(t)]^{a/(1-a)} \exp\left[\frac{b}{1-a} jt\right]. \tag{26}$$

If $P_n(t)$ is given, Eq. 26 establishes the corresponding $P_m(t)$ such that the filter $g(x) = ax + b$ is optimum. If $P_m(t)$ is a characteristic function, it is necessary that $a/(1-a) \geq 0$; hence, $0 \leq a \leq 1$. This condition can be shown by using the following properties of characteristic functions.

$$\left.\begin{array}{l} P(0) = 1 \\[2mm] |P(t)| \leq 1 \qquad \text{all } t \\[2mm] P(-t) = P^*(t) \end{array}\right\} \tag{27}$$

in which $P^*(t)$ is the complex conjugate of $P(t)$.

We proceed with our discussion of Eq. 26 by looking for the classes of messages and noises which satisfy this equation for any value of the noise level. If the noise level is changed, we have

$$p_n(v) \rightarrow 1/c \, p_n(v/c)$$

$$P_n(t_1) \rightarrow P_n(ct_1),$$

in which $c \geq 0$.

We assume that Eq. 26 is satisfied for a specific noise, and for $c = 1$ we have

$$P_m(t_1) = [P_n(t_1)]^{k_1} e^{j\ell_1 t_1},$$

with $k_1 = \dfrac{a}{1-a} \geq 0$ and $b_1 = \dfrac{b}{1-a}$.

If the noise level is changed, the message stays unchanged, and the optimum filter is still of the form $g_1(x) = a_1 x + b_1$, then we have

$$P_m(t_1) = [P_n(ct_1)]^{k_2} e^{j\ell_2 t_1} \qquad k_2 \geq 0. \tag{28}$$

Hence,

$$P_n(t_1) = [P_n(ct_1)]^k e^{j\ell t_1}, \tag{29}$$

in which $k = \dfrac{k_2}{k_1} \geq 0$ and $\ell = \ell_1 - \ell_2$.

We see that if the filter is linear for any value of c, then for any $c > 0$ there is a k and an ℓ such that Eq. 29 holds. If Eq. 29 defines a class of noise characteristic functions, we see by Eq. 28 that the message characteristic function $P_m(t_1)$ belongs

to the same class.

We show in Appendix C that all characteristic functions that satisfy Eq. 29 belong to stable distribution functions. A distribution function $F(x)$ is of a stable type if, for any $a_1 > 0$, b_1, $a_2 > 0$, b_2 there is $a > 0$ and b such that we have

$$F(a_1 x + b_1) * F(a_2 x + b_2) = F(ax+b),$$

where the $*$ indicates composition. Another concise definition is: A type of distribution is stable if it contains all of the compositions of the distributions belonging to it.

Distributions of a stable type have characteristic functions given by

$$P(t) = \exp\{j\gamma t - d|t|^a [1 + j\delta \operatorname{sgn} t \, \omega(t, a)]\}, \tag{30}$$

in which

$$-1 \leq \delta \leq 1 \qquad\qquad 0 < a \leq 2$$

$$d \geq 0 \qquad\qquad \gamma \text{ real}$$

$$\omega(t, a) = \begin{cases} \tan\dfrac{\pi}{2} a & \text{if } a \neq 1 \\[2mm] \dfrac{2}{\pi}\ln |t| & \text{if } a = 1 \end{cases}$$

Only a few of the corresponding probability densities are known. For $a = 2$ we have the Gaussian probability density

$$p(x) = \frac{1}{\sqrt{\pi d}} \exp\left[-\frac{(x-\gamma)^2}{d} \right].$$

For $a = 1$ and $\delta = 0$, we have the Cauchy probability density

$$p(x) = \frac{1}{\pi} \frac{d}{d^2 + (x-\gamma)^2}.$$

Gnedenko and Kolmogorov[11] give the probability density for $a = 1/2$, $\delta = 1$, $\gamma = 0$, and $d = 1$.

$$p(x) = \begin{cases} 0 & x < 0 \\[2mm] \dfrac{1}{\sqrt{2\pi}} e^{-1/2x} x^{-3/2} & x \geq 0 \end{cases}$$

Therefore, the mathematical derivations of this section lead to the following results.

(i) Whenever message and noise have characteristic functions that are related by Eq. 26, a linear filter of the form $g(x) = ax + b$ is optimum.

(ii) If both message and noise have a characteristic function given by Eq. 30, in which a and δ are the same for both message and noise, then the optimum

29

mean-square, no-memory filter is linear, independently of the noise level.

4.2 DISCUSSION OF RESULTS

We recall that we considered the optimum attenuator as a reference system giving no separation of the message from the noise and that we defined a performance index as the ratio of the mean-square error of an arbitrary filter to the mean-square error of the optimum attenuator. We discuss now, in this light, the cases for which the optimum nonlinear no-memory filter reduces to an attenuator, at least at some noise level.

a. No Separation at Any Noise Level

In section 4.1 we have the result that no separation at any noise level will be possible if both message and noise have characteristic functions of a stable type given by Eq. 30. This property of linearity of the optimum mean-square filter is commonly associated only with Gaussian message and Gaussian noise. The reason is that, since the Gaussian probability distribution is the only stable distribution with a finite variance, it will be met in most physically motivated problems. Similarly, since the variance of all other stable distributions is infinite, the concept of an optimum mean-square filter has no meaning for them. It still makes sense, however, to use the conditional mean of the message, given the input as an estimate of the present value of the message. (By referring to Eq. 19, it is easy to see that whether or not the variance (or even the mean) of the message and the noise exists, the conditional mean will exist.)

If the conditional mean is used as an estimate, the consideration of the noise level, which is no longer directly related to the noise variance, is still relevant. To develop further the similarities between all stable distributions and the Gaussian distribution, we give some interesting properties of stable distributions established by Gnedenko and Kolmogorov.[11]

Stable distributions are closely tied to the limit distributions of normalized sums. Consider the random variables, ξ_k, which are independent and have the same distribution function $F(x)$. Consider the normalized sum

$$\zeta_n = \frac{1}{B_n} \sum_{k=1}^{n} \xi_k - A_n. \tag{31}$$

The distribution functions of sums (31) may converge to a limit $V(x)$ for suitably chosen constants A_n and B_n. The following theorem has been established.[12]

THEOREM: In order that the distribution function $V(x)$ be a limit distribution for sums (31) of independent and identically distributed summands, it is necessary and sufficient that it be stable.

If the distribution functions of sums (31) converge to a distribution function $V(x)$ as $n \to \infty$, then we say that $F(x)$, the distribution function of each of the summands, is attracted to $V(x)$. The totality of distribution functions attracted to $V(x)$ is called the

domain of attraction of the distribution function $V(x)$. Only stable distributions have domains of attraction. An important feature is that although the Gaussian distribution attracts a wide class of distributions, the domains of attraction of other stable distributions consist only of those distributions whose character recalls the character of the attracting distribution. More specifically, all distributions with a finite variance will be attracted to the Gaussian distribution.

For other stable distributions we have, for instance, the result of Gnedenko and Kolmogorov,[13] given here for illustration. We consider

$$\lim_{n \to \infty} \text{Prob} \left\{ \frac{1}{a \, n^{1/a}} \sum_{k=1}^{n} \xi_k - A_n < x \right\} = V(x), \tag{32}$$

in which $a > 0$, and a is the characteristic exponent of the stable distribution $V(x)$. Note that a specific choice of B_n in the sum (31) was made to obtain (32). If $F(x)$ is the probability distribution function of any of the ξ and the stable law $V(x)$ with characteristic exponent a, $0 < a < 2$, is the limit distribution in (32), then it is necessary and sufficient that

$$F(x) = \begin{cases} [c_1 a^a + a_1(x)] \dfrac{1}{|x|^a} & x < 0 \\[3mm] 1 - [c_2 a^a + a_2(x)] \dfrac{1}{x^a} & x > 0, \end{cases}$$

in which the functions $a_1(x)$ and $a_2(x)$ satisfy the conditions

$$\lim_{x \to -\infty} a_1(x) = \lim_{x \to +\infty} a_2(x) = 0.$$

b. No Separation at Some Finite and Nonzero Noise Level

We have seen that whenever the characteristic function of the message and the noise are related by

$$P_m(t) = [P_n(ct)]^{a/(1-a)} \exp\left[\frac{b}{1-a} jt\right], \tag{33}$$

in which c is a positive constant and $0 \le a \le 1$, then the optimum no-memory filter reduces to an attenuator. If Eq. 29 is not satisfied at the same time, then this reduction will occur only for a specific noise level related to the value of the constant c. Before discussing these special cases we shall, for the sake of contrast, present some results about the performance index of linear systems with memory.

For linear systems we have seen that no separation is achieved whenever

$$\Phi_{nn}(\omega) = k \, \Phi_{mm}(\omega), \tag{34}$$

in which k is the noise level, and we know that some separation will be obtained if Eq. 34 is not satisfied. We have seen, too, that for any message and noise characteristics the performance index goes to one at zero noise level or at a very large noise level. Hence, two types of behavior are possible for optimum linear systems:

(i) Equation 34 is satisfied and the performance index is equal to one at all noise levels.

(ii) Equation 34 is not satisfied and the performance index is equal to one for $k = 0$ or $k \to \infty$.

A third type of behavior is possible for nonlinear no-memory filters. If Eq. 33 is satisfied but Eq. 29 is not, then the performance index will be equal to one at some finite, nonzero noise level. We shall, however, get separation of the message from the noise at other noise levels. We discuss this point further by considering separately the following two cases.

CASE 1

The message and the noise have the same characteristic function; hence, Eq. 33 is satisfied for $c = 1$, $b = 0$.

No separation will be possible for a noise level equal to 1, and the optimum no-memory mean-square filter will then be an attenuator $g(x) = x/2$. Separation will be possible, however, at other noise levels, and it is clear that the behavior at noise level k will be closely related to the behavior at noise level $1/k$. We shall prove that the performance index $\eta(k)$ is indeed the same at these two noise levels, that is, $\eta(k) = \eta(1/k)$. To show this we make use of the following properties of the optimum nonlinear no-memory filter.

(a) If both message and noise are multiplied by a constant c, then the optimum filter changes from $g(x)$ to $cg(x/c)$, and the error goes from $\overline{e^2}$ to $c^2\overline{e^2}$, or, in tabular form,

Message	Noise	Filter	Error
$p_m(v)$	$p_n(v)$	$g(x)$	$\overline{e^2}$
$\frac{1}{c}p_m(v/c)$	$\frac{1}{c}p_n(v/c)$	$cg(x/c)$	$c^2\overline{e^2}$

This property can be easily established by direct substitution in the expressions for the optimum mean-square filter and for the resulting error.

(b) If the characteristics of the message and the noise are interchanged, then the filter changes from $g(x)$ to $x - g(x)$, and the error $\overline{e^2}$ stays the same. In tabular form we have

Message	Noise	Filter	Error
$p_m(v)$	$p_n(v)$	$g(x)$	$\overline{e^2}$
$p_n(v)$	$p_m(v)$	$x-g(x)$	$\overline{e^2}$

32

This property was established by Schetzen.[14]

Now we consider the filtering of a signal consisting of a message with probability density $p(v)$ added to noise with a probability density that is either $\frac{1}{c}p(v/c)$ or $cp(cv)$.

We recall that the performance index is $\eta(k) = \frac{\overline{e^2}}{\overline{m^2}} \frac{1+k}{k}$, in which $k = \frac{\overline{n^2}}{\overline{m^2}}$, and thus we assume that the means of both message and noise are zero. Here, if $p_m(v) = p(v)$ and $p_n(v) = \frac{1}{c}p(v/c)$, then $k = c^2$.

We have Table 1 from properties a and b.

Table 1.   Effect of noise level on the optimum filter.

	Message	Noise	Noise Level	Filter	Error	Performance Index
by definition	$p(v)$	$\frac{1}{c}p(v/c)$	k	$g(x)$	$\overline{e^2}$	$\frac{\overline{e^2}}{\overline{m^2}} \frac{1+k}{k} = \eta$
by property b	$\frac{1}{c}p(v/c)$	$p(v)$	$1/k$	$x-g(x)$	$\overline{e^2}$	$\frac{\overline{e^2}}{k\overline{m^2}} \frac{1+\frac{1}{k}}{1/k} = \eta$
by property a	$p(v)$	$cp(cv)$	$1/k$	$x-\frac{1}{c}g(cx)$	$\frac{\overline{e^2}}{k}$	$\frac{\overline{e^2}/k}{\overline{m^2}} \frac{1+\frac{1}{k}}{1/k} = \eta$

Hence, we have established that $\eta(k) = \eta(1/k)$ whenever the message and the noise have the same characteristic function. Since $\eta(0) = 1$, we have $\eta(k) \to 1$ as $k \to \infty$; hence, the optimum nonlinear no-memory filter reduces to an attenuator for $k \to \infty$.

EXAMPLE: We consider a signal made of message and additive noise with the probability densities shown in Fig. 12, in which k is the noise level. For this signal, the optimum nonlinear no-memory filter is odd and for $x \geq 0$ is given by
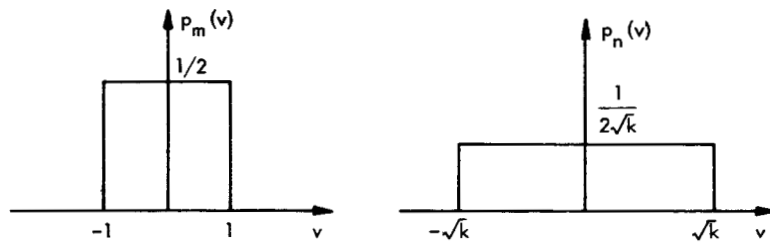
Fig. 12.   Example of message and noise with the same characteristic functions.

33
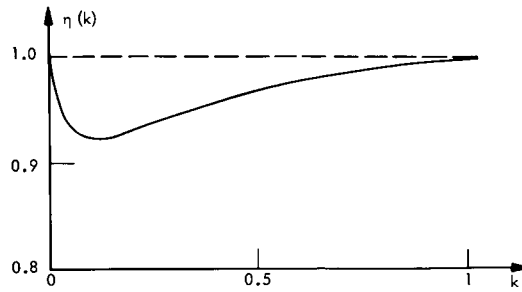
Fig. 13. Performance index for the example of Fig. 12.

$$g(x) = \begin{cases} x & 0 \leq x \leq 1 - \sqrt{k} \\ \frac{1}{2}(1+x-k) & 1 - \sqrt{k} \leq x \leq 1 + \sqrt{k} \\ \text{Not defined} & x > 1 + \sqrt{k} \end{cases}$$

We have assumed that $k \leq 1$. If $k > 1$ the optimum filter $g_1(x)$ is given by

$$g_1(x) = x - \frac{1}{\sqrt{k}} g(\sqrt{k} \, x),$$

in which $g(x)$ is the optimum filter for the noise level $k_1 = \frac{1}{k} < 1$. The minimum normalized mean-square error is

$$e^2 = k\left(1 - \frac{\sqrt{k}}{2}\right) \qquad k \leq 1,$$

and the performance index becomes

$$\eta(k) = (1+k)\left(1 - \frac{\sqrt{k}}{2}\right) \qquad k \leq 1.$$

For $k > 1$, $\eta(k) = \eta(1/k)$. The performance index as a function of the noise level is given in Fig. 13.

CASE 2

The message and the noise have different characteristic functions, but Eq. 33 is satisfied.

There are numerous such cases, both discrete and continuous, for which no separation will occur.

(a) We consider the Poisson distribution as a specific example of discrete message and discrete noise. Let the message have the distribution

$$P_m(x=k) = \frac{\lambda_m^k}{k!} e^{-\lambda_m} \qquad\qquad P_m(t) = \exp\left[\lambda_m(e^{jt}-1)\right]$$

and, similarly, for the noise the distribution is

34

$$P_n(x=k) = \frac{\lambda_n^k}{k!} e^{-\lambda_n} \qquad P_n(t) = \exp\left[\lambda_n(e^{jt}-1)\right].$$

Then the input signal has the distribution

$$P_{m+n}(x=k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad P_{m+n}(t) = \exp[\lambda(e^{jt}-1)]$$

in which $\lambda = \lambda_m + \lambda_n$.
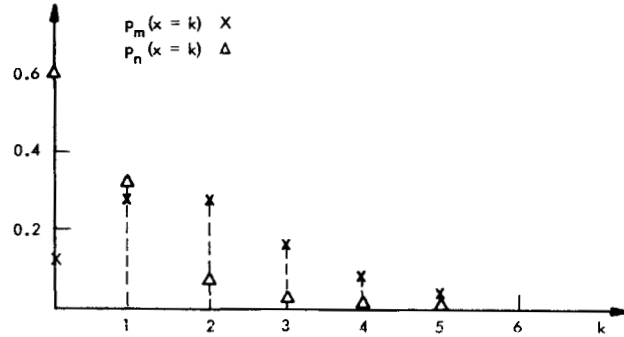
Fig. 14. Poisson distributions for which the filter is linear.

Since Eq. 33 is satisfied, the optimum nonlinear no-memory filter is an attenuator $g(x) = ax$, and the attenuation constant a is given by

$$a = \frac{\lambda_m}{\lambda_m + \lambda_n}.$$

The probability distribution of the message and the noise, however, can be quite different. For example, we take $\lambda_m = 2$ and $\lambda_n = 1/2$; then, $g(x) = 0.8\,x$. The graphs for the probability densities of the message and the noise, given in Fig. 14, show them to be quite different. Note that both message and noise have Poisson distributions, but that changing the value of λ here brings a change in the shape of the distribution functions and, hence, is not equivalent to a change of level.

(b) For continuous distributions, a noteworthy class of messages and noises for which there is no separation at some noise level is characterized by a probability density of the Γ type for which the characteristic function is given by

$$P(t) = [1-jct]^{-q},$$

in which q is a positive constant. To this class belong the exponential probability density

$$p(v) = \frac{1}{c} e^{-v/c} u_{-1}(v) \qquad \text{for } q = 1$$

35

and the probability density

$$p(v) = \frac{v}{c^2} e^{-v/c} u_{-1}(v) \qquad \text{for } q = 2.$$

For these probability densities it is easy to show that the mean m and the variance σ^2 are given by

$$m = cq$$

$$\sigma^2 = c^2 q.$$

Assume that the characteristic functions of the message and the noise are given by

$$P_m(t) = [1-jc_m t]^{-q_m}$$

$$P_n(t) = [1-jc_n t]^{-q_n}.$$

Then Eq. 33 in the form $P_m(t) = [P_n(c_n t)]^{a/1-a}$ will be satisfied whenever $c_m = c_n$, or, in terms of the noise level, whenever

$$k = \frac{\sigma_n^2}{\sigma_m^2} = \frac{q_n}{q_m}.$$

For example, we consider

$$P_m(t) = [1-jt]^{-1} \qquad\qquad p_m(v) = e^{-v} u_{-1}(v)$$

$$P_n(t) = [1-jct]^{-2} \qquad\qquad p_n(v) = \frac{v}{c^2} e^{-v/c} u_{-1}(v).$$

(35)

The optimum nonlinear no-memory filter $g(x)$, when c is used as a parameter, is given by

$$g(x) = \beta \, \frac{e^{-\beta x}[\beta x+2] + \beta x - 2}{e^{-\beta x} + \beta x - 1},$$

in which $\beta = 1 - \frac{1}{c}$.

Whenever $c = 1$ or $k = 2$, the optimum filter reduces to an attenuator, as illustrated graphically in Fig. 15, which gives $g(x)$ for $k = 1/2$, 2, and 8.

c. No Separation at Some Noise Level and Infinitely Divisible Distributions

Cases 1 and 2 have the common characteristic that both the message and the noise can be considered as the sum of n_1 and n_2 independent random variables with the same probability distribution. If η and ζ are the amplitude random variables of message and
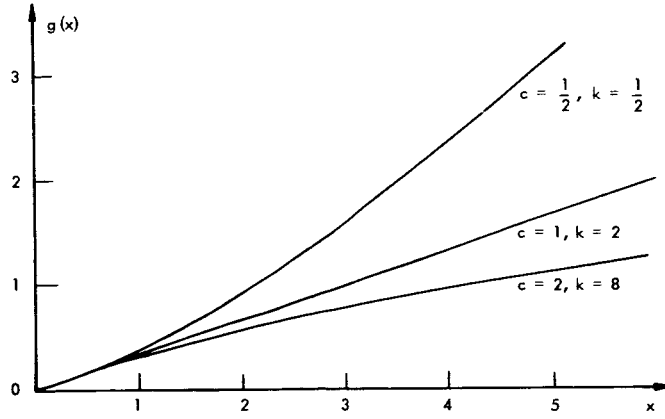
Fig. 15. Filter for message and noise of Eqs. 35.

noise and $n_1, n_2 = 1, 2, \ldots$, then

$$\eta = \sum_{k=1}^{n_1} \xi_{mk}$$

$$\zeta = \sum_{j=1}^{n_2} \xi_{nj}$$

in which all of the ξ_m and ξ_n are independent random variables, each associated with the same characteristic function $P(t)$. It is clear that we have

$$P_m(t) = [P(t)]^{n_1}$$

$$P_n(t) = [P(t)]^{n_2}$$

and, therefore,

$$P_m(t) = [P_n(t)]^{n_1/n_2} \tag{36}$$

which is a special case of Eq. 33. This fact allows us to form a great number of characteristic functions for which Eq. 33 is satisfied. Note that the exponent $a/(1-a) = n_1/n_2$ of Eq. 33 is then rational. Cases for which $a/(1-a)$ is not constrained to be rational can be found by considering infinitely divisible distributions.[15]

DEFINITION: The random variable ξ is infinitely divisible if for every number n it can be represented as the sum $\xi = \sum_{j=1}^{n} \xi_{nj}$ of n independent identically random variables $\xi_{n_1}, \xi_{n_2} \cdots \xi_{n_n}$.

37

The property of infinitely divisible distributions of interest here is given by the following theorem.

THEOREM: If $P(t)$ is the characteristic function of an infinitely divisible distribution, then, for every $c > 0$, $[P(t)]^c$ is also a characteristic function.

If either message or noise has an infinitely divisible distribution, then, for any $a/(1-a) = c > 0$, we can find a noise (or message) such that the filter $g(x) = ax$ is optimum.

The Poisson, Cauchy, Gaussian, and Γ type of distributions are infinitely divisible. In fact, stable distributions represent a subclass of infinitely divisible distributions.

4.3 EXTENSION TO NON MEAN-SQUARE FILTERING

The results of Section III on the equivalence of mean-square and non mean-square filtering allow some extension of the results of this section to non mean-square filtering. These extensions are based on properties of the conditional probability density, Eq. 19.

a. No Separation at Any Noise Level

Among the probability densities of a stable type for which we have an explicit expression, the Gaussian is the only one leading to a conditional probability density that is even. Hence, only for Gaussian message and noise can we say that the optimum filter will be linear for all criteria discussed in Section III. For all other probability densities for which the conditional mean is linear, we can state only that the mean-square and the non mean-square filters will be different.

b. No Separation at Some Noise Level

We shall prove, first, that when message and noise have the same probability density, an attenuator is the optimum filter for most error criteria. Then, by use of a counter-example, we show that if message and noise have different probability densities, the results obtained for mean-square filtering do not extend to non mean-square filtering.

(i) It is easy to verify that, whenever message and noise have the same probability density, then the conditional probability density $p_{m/m+n}(v/x)$ is even about the point $v = x/2$. This allows us to say that the optimum filter will be $g(x) = x/2$ for all error-weighting functions considered in Section III, with the exception of criterion d. If some further property of the probability density of message and noise leads, for instance, to a unimodal conditional probability density, then the filter will be $g(x) = x/2$ for criterion d also. To illustrate this point further we consider the following example.

EXAMPLE: The probability density

$$p_m(v) = p_n(v) = \begin{cases} |v| & |v| < 1 \\ 0 & |v| \geq 1 \end{cases}$$

and the corresponding conditional probability density for some x are shown in Fig. 16. For most criteria $g(x) = x/2$ is the optimum filter. For criterion d (Section III) we
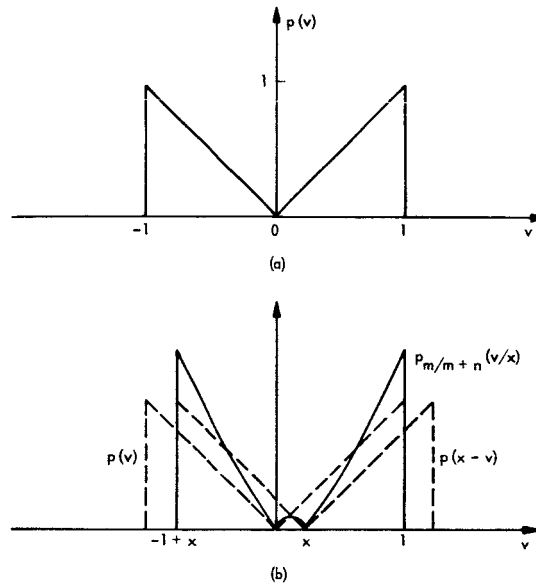
Fig. 16. Probability density for which the filter is not unique.

take

$$W(e) = \begin{cases} 0 & |e| < 0.1 \\ 1 & |e| \geqslant 0.1. \end{cases}$$

For the value of x used in Fig. 16b, either the value $g(x) = 0.9$ or $g(x) = x - 0.9$, symmetric with respect to $x/2$, is optimum. For larger values of x, $g(x) = x/2$ will be optimum. The optimum filter characteristic $g(x)$ is shown in Fig. 17. For $|x| < |x_o|$ the filter is not unique and is given by the dotted line or by the solid line in Fig. 17. For $|x| \geqslant |x_o|$ we have a unique filter $g(x) = x/2$.

This result can be generalized and extended to cases with memory by making use of
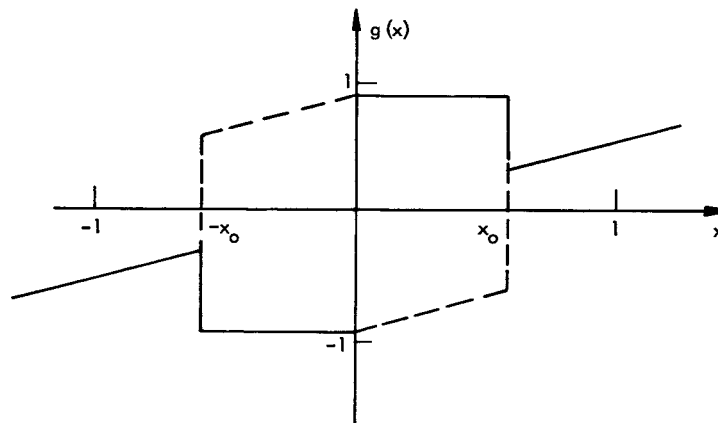
Fig. 17. Filter for the probability density of Fig. 16.

39

a result established by Schetzen.[14] He showed that if for an even error-weighting function the characteristics of message and noise are interchanged, then the optimum functional of the input, $\mathcal{K}_o[x(t)]$, goes to

$$\mathcal{K}_1[x(t)] = x(t) - \mathcal{K}_o[x(t)].$$

If the characteristics of message and noise are the same and the optimum functional is unique, then we obviously have

$$\mathcal{K}_o[x(t)] = \mathcal{K}_1[x(t)] = \frac{x}{2}.$$

If the optimum functional is not unique and if $\mathcal{K}_o[x(t)]$ is an optimum functional, then another one will be

$$\mathcal{K}_1[x(t)] = x(t) - \mathcal{K}_o[x(t)].$$

(ii) For message and noise having different probability densities, we show by an example that if the optimum nonlinear mean-square filter is linear, the optimum non mean-square filter need not be.

EXAMPLE: Consider the message and noise probability densities given in Fig. 18.
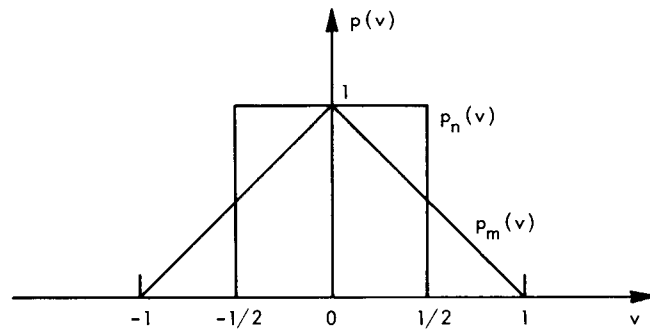
Fig. 18.   Example for which the mean-square filter is linear but the
absolute value of the error filter is nonlinear.

This example has been chosen to give $g_1(x) = 2x/3$ for the optimum no-memory mean-square filter. The optimum no-memory filter for the absolute value of the error criterion can be shown to be $g_2(x)$, odd, and to have for $x \geq 0$ the expression

$$g_2(x) = \begin{cases} 1 - \sqrt{1-x} & 0 \leq x \leq \frac{1}{2} \\ \\ \frac{1}{\sqrt{2}}\left[x - \frac{3}{2}\right] + 1 & x \geq \frac{1}{2}. \end{cases} \tag{37}$$

It is clear that this second filter is nonlinear.

4.4 EXTENSION TO MEAN-SQUARE FILTERING WITH MEMORY

If we now restrict our attention to mean-square filtering, some extension of the results for no-memory filters to filters with memory is possible.

We first prove that: If we let the message $m(t)$ be characterized by some statistics and assume that the noise $n(t)$ can be considered as the sum of two statistically independent processes

$$n(t) = n_1(t) + n_2(t),$$

and let $n_1(t)$ and $n_2(t)$ have statistics identical to the statistics of the message, then, if

$$x(t) = m(t) + n(t)$$

is the present value of the input, the optimum mean-square filter is an attenuator with

$$\mathcal{W}_{opt}[x(t)] = \frac{x(t)}{3}. \tag{38}$$

PROOF: The optimum mean-square filter is the conditional mean.

$$\mathcal{K}_{opt}[x(t)] = E[m(t)/x(t_1), \ t_1 \leqslant t].$$

Now, by symmetry, we have

$$E[m(t)/x(t_1), \ t_1 \leqslant t] = E[n_1(t)/x(t_1), \ t_1 \leqslant t] = E[n_2(t)/x(t_1), \ t_1 \leqslant t].$$

Furthermore,

$$E[n_1(t)+n_2(t)/x(t_1), \ t_1 \leqslant t] = 2 \ E[n_1(t)/x(t_1), \ t_1 \leqslant t] = 2 \ \mathcal{K}_{opt}[x(t)]$$

and

$$E[n_1(t)+n_2(t)/x(t_1), \ t_1 \leqslant t] = E[x(t)-m(t)/x(t_1), \ t_1 \leqslant t] = x(t) - \mathcal{K}_{opt}[x(t)].$$

Therefore, Eq. 38 holds.   Q. E. D.

It is clear that this result generalizes for

$$m(t) = \sum_{k=1}^{n_1} m_k(t)$$

$$n(t) = \sum_{j=1}^{n_2} n_j(t),$$

in which all m_k and n_j are statistically independent processes with identical statistics. In such cases the optimum mean-square filter is an attenuator.

$$\mathcal{H}_{opt}[x(t)] = \frac{n_1}{n_1 + n_2} \, x(t).$$

Note that this result will hold only at a specific noise level.

V. OTHER PROPERTIES OF NONLINEAR NO-MEMORY FILTERS
IN THE MEAN-SQUARE SENSE

In this section we make use of a result of Section IV to establish some further pro-
perties of nonlinear no-memory filters in the mean-square sense. We first consider
the characterization of the message and the noise for which the optimum mean-square
filter is of a given form. As a result of this investigation we derive simple expressions
for the mean-square error when the noise is either Gaussian or Poisson. We take advan-
tage of the error expression for Gaussian noise to find the message probability density
that gives the largest mean-square error in an optimum nonlinear no-memory filter
under the constraint that the average message power be constant. The resulting mes-
sage probability density is Gaussian.

The result of Section IV that we shall need is the expression for the optimum mean-
square filter in terms of the characteristic functions of the message and the noise:

$$g(x) = \frac{\int P_n(t) \frac{dP_m(t)}{djt} e^{-jtx} dt}{\int P_n(t) P_m(t) e^{-jtx} dt} = \frac{r(x)}{q(x)}. \tag{39}$$

5.1 OPTIMUM NONLINEAR NO-MEMORY FILTER OF A PRESCRIBED FORM

This problem has been considered quite generally by Balakrishnan.[10] For mean-
square filtering and no-memory filters, his approach leads to a differential equation
relating the characteristic functions of the message and the noise whenever the filter
$g(x)$ has the form of a specified polynomial in x. We shall extend this result to the
case in which $g(x)$ is a ratio of polynomials.

Recently, Tung[16] obtained the probability density of the input of the optimum filter
as a function of the filter characteristic $g(x)$ when the noise is Gaussian. We shall
establish necessary relations between the input density and the filter characteristic
when the noise is Poisson or of the Γ type.

a. Filter Characteristic g(x), a Ratio of Polynomials

We have

$$g(x) = \frac{\sum_{k=0}^{M} a_k x^k}{\sum_{\ell=0}^{N} b_\ell x^\ell},$$

and therefore Eq. 39 becomes

$$\left[\sum_{k=0}^{M} a_k x^k \right] q(x) = \left[\sum_{\ell=0}^{N} b_\ell x^\ell \right] r(x) .$$

By equating the Fourier transforms, we have

$$\sum_{k=0}^{M} a_k \frac{d^k [P_n(t) P_m(t)]}{[d(jt)]^k} = \sum_{\ell=0}^{N} b_\ell \frac{d^\ell \left[P_n(t) \dfrac{dP_m(t)}{d(jt)} \right]}{[d(jt)]^\ell} .$$

If we know the characteristic function of the noise, this is a linear differential equation with variable coefficients for the characteristic function of the message.

b. Relation between the Input Density Function and the Filter Characteristic for Various Types of Noise

Consider again Eq. 39, which, by writing r(x) only in terms of its Fourier transform, may be written

$$g(x) \, q(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} P_n(t) \frac{dP_m(t)}{d(jt)} e^{-jtx} dt, \tag{40}$$

but we can write

$$P_n(t) \frac{dP_m(t)}{d(jt)} = \frac{d}{d(jt)} [P_n(t) P_m(t)] - P_m(t) \frac{dP_n(t)}{d(jt)} .$$

Since q(x) is the Fourier transform of $P_n(t) P_m(t)$, we have

$$[g(x) - x] \, q(x) = -\frac{1}{2\pi} \int_{-\infty}^{+\infty} P_m(t) \frac{dP_n(t)}{d(jt)} e^{-jtx} dt . \tag{41}$$

Comparison of (40) and (41) shows that if the characteristics of message and noise are interchanged, then the characteristic of the optimum filter is changed from g(x) to x − g(x). This is a general result holding for filters with memory, as well as non mean-square criteria (see Schetzen[14] and Section IV).

We define F(t) as

$$\frac{dP_n(t)}{d(jt)} = F(t) P_n(t)$$

or

$$F(t) = \frac{d[\ln P_n(t)]}{d(jt)} .$$

Let f(x) be the Fourier transform of F(t). Then by transforming the right-hand side of Eq. 41, we obtain

$$[g(x)-x]\, q(x) = -\int_{-\infty}^{+\infty} q(x_1)\, f(x-x_1)\, dx_1. \tag{42}$$

We shall now apply Eq. 42 to various types of noise. Since (42) is a homogeneous equation, $q(x) \equiv 0$ is a solution. This solution is of no interest here, and henceforth we shall assume that $q(x) \neq 0$.

For Gaussian noise,

$$P_n(t) = e^{-\frac{\sigma^2}{2} t^2}$$

and

$$\frac{d[\ln P_n(t)]}{d(jt)} = F(t) = -\frac{\sigma^2 t}{j}\ .$$

Hence

$$f(x) = -\sigma^2\, u_1(x),$$

in which $u_1(x)$ is the unit doublet occurring at $x = 0$. Equation 42 now takes the form

$$[g(x)-x]\, q(x) = \sigma^2\, \frac{dq(x)}{dx},$$

which can be integrated to give

$$q(x) = \exp\left[-\frac{1}{\sigma^2}\int_{-\infty}^{x} [v-g(v)]\, dv\right].$$

This is the expression obtained by Tung.[16]

For Poisson noise,

$$P_n(t) = e^{\lambda(e^{jt}-1)}$$

and

$$\frac{d[\ln P_n(t)]}{d(jt)} = F(t) = \lambda\, e^{jt}.$$

Therefore

$$f(x) = \lambda\, u(x-1)\ .$$

We denote by $u(x-1)$ the unit impulse occurring at $x = 1$. Equation 42 now takes the form

$$[g(x)-x]\, q(x) = -\lambda\, q(x-1),$$

which can be written

45

$$q(x+1) - A(x) \, q(x) = 0. \tag{43}$$

Here, we let

$$A(x) = \frac{\lambda}{x + 1 - g(x+1)} \, .$$

The solution of the difference equation (43) is well known.[17]

Let $t(x)$ be an arbitrary single-valued function defined in a unit interval $a \leqslant x < a+1$. Then we have

$$q(x) = t(a_x) A(a_x) A(a_x+1) \ldots A(x-1) \qquad x \geqslant a,$$

in which a_x is the point in the interval $a \leqslant x < a+1$ which is such that $x - a_x$ is an integer.

For noise of the Γ type,

$$P_n(t) = [1-cjt]^{-\gamma} \qquad \gamma, \, c > 0$$

Hence

$$\frac{d[\ln P_n(t)]}{d(jt)} = F(t) = c\gamma[1-cjt]^{-1}$$

and

$$f(x) = \begin{cases} \gamma \, e^{-x/c} & x \geqslant 0 \\ \\ 0 & x < 0 \, . \end{cases}$$

Now Eq. 42 takes the form

$$[g(x)-x] \, q(x) = - \int_{-\infty}^{x} q(x_1) \, \gamma \, e^{-(x-x_1)/c} \, dx_1 \, ,$$

which can be written

$$[x-g(x)] \, q(x) \, \frac{e^{x/c}}{\gamma} = \int_{-\infty}^{x} q(x_1) \, e^{x_1/c} \, dx_1 \, .$$

By differentiation,

$$\frac{d}{dx} \left\{ [x-g(x)] \, q(x) \, \frac{e^{x/c}}{\gamma} \right\} = q(x) \, e^{x/c} \, .$$

This is a differential equation for $g(x)$ that can be written

$$\frac{dq(x)}{dx} \left\{ \frac{x-g(x)}{\gamma} \right\} + q(x) \left\{ \frac{1-g'(x)}{\gamma} + \frac{x-g(x)}{\gamma c} - 1 \right\} = 0 \, ,$$

in which

$$g'(x) = \frac{dg(x)}{dx} \, .$$

For $q(x) \neq 0$, we have

$$\frac{\frac{dq(x)}{dx}}{q(x)} = -\frac{1-g'(x)}{x-g'(x)} - \frac{1}{c} + \frac{\gamma}{x-g(x)} ,$$

and the expression for $q(x)$ is

$$q(x) = \frac{B e^{-\frac{x}{c}}}{|x-g(x)|} \exp \left[\int \frac{\gamma \, dx}{x-g(x)} \right] ,$$

in which B is a positive constant.

If $g(x)$, the characteristic of the nonlinear filter, is given and if the noise is of a type considered above, then $q(x)$, the probability density of the input, has to fulfill the following conditions:

(i) $q(x)$ has to satisfy the relation obtained in terms of $g(x)$.

(ii) $q(x)$ has to be a proper density function.

(iii) $P(t) = \frac{Q(t)}{P_n(t)}$ has to be a characteristic function, in which we let $Q(t)$ be the characteristic function for $q(x)$, and $P_n(t)$ be the characteristic function for the noise. The third requirement comes from the fact that

$$Q(t) = P_m(t) \, P_n(t),$$

in which $P_m(t)$ is the characteristic function of the message.

5.2 EXPRESSION FOR THE MEAN-SQUARE ERROR

In Eq. 42 we can express $g(x)$ in terms of $q(x)$ by writing

$$g(x) = \frac{-\int q(x_1) \, f(x-x_1) \, dx_1}{q(x)} + x.$$

Whenever $f(x)$ is a singularity function, this relation for $g(x)$ leads to a simple expression for the mean-square error in terms of $q(x)$, the input probability density. The two well-known, nontrivial noise characteristics that give a singularity function for $f(x)$ are Gaussian noise and Poisson noise. We shall use the following expression for the mean-square error:

$$\overline{e^2} = \overline{m^2} - \int g^2(x) \, q(x) \, dx,$$

in which $\overline{m^2}$ is the mean-square value of the message. This expression is obtained without difficulty by using the known result that the error resulting from the optimum mean-square filter is uncorrelated with the output of all nonlinear no-memory filters

47

with the same input.

a. Poisson Noise

$$P(x=k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$P(t) = e^{\lambda(e^{jt}-1)}$$

hence

$$f(x) = \lambda u(x-1),$$

and

$$g(x) = -\lambda \frac{q(x-1)}{q(x)} + x.$$

$$\overline{e^2} = m^2 - \int \left[-\lambda \frac{q(x-1)}{q(x)} + x\right]^2 q(x)\ dx$$

$$= m^2 - \int x^2 q(x)\ dx + 2\lambda \int xq(x-1)\ dx - \lambda^2 \int \frac{q^2(x-1)}{q(x)}\ dx.$$

Since

$$\int xq(x-1)\ dx = \int (x+1)q(x)\ dx = \overline{m} + \lambda + 1$$

and

$$\int x^2 q(x)\ dx = \overline{m^2} + \lambda + \lambda^2 + 2\overline{m}\lambda,$$

then we have

$$\overline{e^2} = \lambda + \lambda^2 - \lambda^2 \int \frac{q^2(x-1)}{q(x)}\ dx. \tag{44}$$

Equation 44 holds whether or not the message has zero mean.

b. Gaussian Noise

$$p_n(x) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

$$P_n(t) = \exp\left(-\frac{\sigma^2 t^2}{2}\right),$$

hence

48

$$f(x) = -\sigma^2 u_1(x)$$

and

$$g(x) = \sigma^2 \frac{q'(x)}{q(x)} + x.$$

$$\overline{e^2} = m^2 - \int \left[\sigma^2 \frac{q'(x)}{q(x)} + x\right]^2 q(x)\, dx$$

Since

$$\int x^2 q(x)\, dx = \overline{m^2} + \sigma^2$$

and

$$\int xq'(x)\, dx = xq(x)\Big|_{-\infty}^{+\infty} - \int q(x)\, dx = -1,$$

then we have

$$\overline{e^2} = \sigma^2 - \sigma^4 \int \frac{q'^2(x)}{q(x)}\, dx. \tag{45}$$

5.3 MAXIMUM OF THE ERROR UNDER CONSTRAINTS FOR ADDITIVE GAUSSIAN NOISE

Since we have an expression for the mean-square error solely in terms of the input probability density, we can find extrema of the error under constraint by the method of calculus of variations. We shall consider a power constraint on the input.

We consider here a filtering problem characterized by additive Gaussian noise of known average power. We consider all possible messages of fixed average power, and in each case use the optimum nonlinear no-memory filter in the mean-square sense to separate the message from the noise. We now undertake to find the message probability density that gives an extremum of the mean-square error. Since the message and the noise are statistically independent, a constraint on the input average power is equivalent to a constraint on the message average power, and we write

$$\int x^2 q(x)\, dx = \overline{m^2} + \sigma^2.$$

Other constraints are

$$\int q(x)\, dx = 1$$

$$q(x) \geq 0 \qquad \text{for all } x.$$

We take care of the last constraint by letting

$$y^2(x) \overset{\triangle}{=} q(x),$$

and we have

$$\overline{e^2} = \sigma^2 - 4\sigma^4 \int y'^2(x) \, dx.$$

Because $q(x)$ is the result of convolving a Gaussian probability density with the message probability density $p_m(x)$, we would need another constraint on $q(x)$ to ensure that $p_m(x)$ is positive. This constraint cannot be handled analytically, however, and we shall have to select among the solutions obtained for $q(x)$ those leading to an acceptable probability density. In terms of $y(x)$, using the Lagrange multiplier, we look for extrema of

$$J = \int \left[y'^2 + \lambda_1 x^2 y^2 + \lambda_2 y^2 \right] dx, \tag{46}$$

in which λ_1 and λ_2 are Lagrange multipliers. This leads to the following Euler-Lagrange equation:

$$y'' + y \left[\lambda_1 x^2 + \lambda_2 \right] = 0. \tag{47}$$

We have obtained here the Weber-Hermite differential equation. Since we are looking for solutions that are square integrable, we have the boundary conditions

$$y \to 0 \quad \text{for} \quad |x| \to \infty.$$

The differential equation has solutions that satisfy these boundary values[18] only if it is in the form

$$\frac{d^2 y}{du^2} + y \left[n + \frac{1}{2} - \frac{u^2}{2} \right] = 0 \tag{48}$$

in which n, a non-negative integer, is the eigenvalue. The corresponding solutions or eigenfunctions are the Hermite functions

$$y_n(u) = D_n(u) \overset{\triangle}{=} \exp\left(-\frac{u^2}{2} \right) 2^{-n/2} H_n\left(\frac{u}{\sqrt{2}} \right)$$

in which $H_n(v)$ is the Hermite polynomial.

$$H_n(v) \overset{\triangle}{=} (-1)^n e^{v^2} \frac{d^n e^{-v^2}}{dv^n} .$$

To put Eq. 47 in the form of Eq. 48, we let $x = cu$, in which c is a constant, and

thus obtain the solution

$$y_n(x) = A \, D_n\left(\frac{x}{c}\right).$$

Here, A, an arbitrary constant, appears because the linear differential equation to be satisfied is an homogeneous equation. The solution for the amplitude probability density of the input becomes

$$q_n(x) = A^2 D_n^2\left(\frac{x}{c}\right).$$

It can be shown that the minimum of the integral $\int y'^2 \, dx$ that appears with a minus sign in the expression for the mean-square error (Eq. 45) corresponds to the eigenvalue $n = 0$. For $n = 0$ we have

$$q(x) = A^2 \exp\left(-\frac{x^2}{2c^2}\right)$$

which is, therefore, the amplitude probability of the input giving the maximum mean-square error.

We satisfy the constraints by letting $A^2 = 1/\sqrt{2\pi} \, c$, and $c^2 = \sigma^2 + \overline{m^2}$. Therefore,

$$q(x) = \frac{1}{\sqrt{2\pi} \, \sqrt{\sigma^2 + \overline{m^2}}} \exp\left[-\frac{x^2}{2(\sigma^2 + \overline{m^2})}\right].$$

The probability density of the message now is

$$p_m(x) = \frac{1}{\sqrt{2\pi \overline{m^2}}} \exp\left[-\frac{x^2}{2\overline{m^2}}\right].$$

Hence, when the noise is Gaussian and additive, and the message has a fixed average power, the maximum mean-square error is obtained whenever the message is also Gaussian. In such a case, the optimum no-memory filter reduces to an attenuator and

$$\overline{e^2} = \sigma^2 - \sigma^4 \int \frac{q'^2(x)}{q(x)} \, dx = \frac{\sigma^2 \overline{m^2}}{\sigma^2 + \overline{m^2}}.$$

One might wonder if some interpretation can be given in the context to higher order eigenvalues and eigenfunctions ($n = 1, 2$, etc.) which correspond to stationary values of the expression for mean-square error.

However, although $(q(x) = A^2 D_n^2(x/c))$, the probability density of the input, is positive for all x, the corresponding message probability density $p_m(x)$ is not strictly positive for $n > 0$ and does not correspond to a physical situation.

51

VI. PERFECT SEPARATION BY A NONLINEAR NO-MEMORY FILTER

In this section we establish the conditions to be satisfied by the amplitude probability densities of the message and the noise if the message is to be perfectly separated from the noise by a nonlinear no-memory filter. Since for most noises of interest this leads to a trivial message, we have the possibility of finding meaningful lower bounds on the average weighted error. Constraints on the message are proposed which lead to a meaningful set of conditions.

6.1 CONDITIONS FOR PERFECT SEPARATION

We consider a nonlinear no-memory filter with characteristic $g(x)$ and error-weighting function $W(e)$. We require that $W(e)$ be a non-negative function of the error which goes to zero if and only if the error goes to zero.

We consider the expression for the average weighted error

$$\overline{W(e)} = \iint W[v-g(x)]\, p_{m,\,m+n}(v,x)\, dv\, dx \tag{49}$$

$$\overline{W_x(e)} = \int W[v-g(x)]\, p_{m/m+n}(v/x)\, dv. \tag{50}$$

Then we can write

$$\overline{W(e)} = \int \overline{W_x(e)}\, p_{m+n}(x)\, dx.$$

We have $p_{m+n}(x) \geq 0$ for all x and $\overline{W_x(e)} \geq 0$, since $W(e) \geq 0$ and $p_{m/m+n}(v/x) \geq 0$; therefore, we have $\overline{W(e)} = 0$ if and only if $\overline{W_x(e)}\, p_{m+n}(x) = 0$ for all x.

This condition requires either that $\overline{W_x(e)} \neq 0$ only for x such that $p_{m+n}(x) = 0$ or that $p_{m+n}(x) \neq 0$ only for x such that $\overline{W_x(e)} = 0$.

However, $\overline{W_x(e)}$, which is the average weighted error conditioned on the occurrence of the input x, is defined only for inputs that can be obtained. Hence, the only meaningful requirements are $p_{m+n}(x) \neq 0$ and $\overline{W_x(e)} = 0$. From the expression of $\overline{W_x(e)}$ of Eq. 50, therefore, for all x of interest (that is, such that $p_{m+n}(x) \neq 0$), we now require that $W[v-g(x)]\, p_{m/m+n}(v/x) = 0$.

Since $W[v-g(x)]$ is always positive except for $v = g(x)$, we see that $\overline{W_x(e)}$ will be zero if and only if $p_{m/m+n}(v/x)$ consists of a single impulse occurring at $v = g(x)$. Hence, we require that

$$p_{m/m+n}(v/x) = u[v-g(x)]. \tag{51}$$

It is clear that $g(x)$ has to be a single-valued function of x. In terms of p_n and p_m

Eq. 51 becomes

$$\frac{p_n(x-v) \, p_m(v)}{p_{m+n}(x)} = u[v-g(x)].$$

(52)

The left-hand side of (52) is a function of v defined for all x such that $p_{m+n}(x) \neq 0$, and condition (52) states that this function should be a single impulse occurring at $v = g(x)$ for all such x.

Let us take $p_n(v)$ to be the probability density of continuous noise. We have

$$p_{n+m}(x) = \int p_n(x-v) \, p_m(v) \, dv$$

and, since $p_n(v)$ contains no impulses, $p_{m+n}(v)$ contains no impulses. Then, if condition (52) is to be satisfied, $p_m(v)$ has to contain only impulses, and we can make the following statements.

STATEMENT 1: If $p_n(v) > 0$ for all v, then condition (52) is satisfied for any x if and only if $p_m(v) = u(v-v_o)$. Here, v_o denotes a constant.

STATEMENT 2: If $p_n(v) > 0$ for all $v > v_o$, then condition (52) is satisfied for all x such that $p_{m+n}(x) > 0$ if and only if $p_m(v) = u(v-v_1)$. Here, v_o and v_1 are constants.

PROOF OF STATEMENT 2: Take $p_m(v) = u(v-v_1)$; then condition (52) is obviously satisfied by taking $g(x) = v_1$. Conversely, assume that $p_m(v) = au(v-v_1) + f_1(v)$, in which $0 < a < 1$ and $f_1(v) > 0$ for $v_2 \leq v \leq v_3$; then for $x \geq$ (the larger of v_o, v_3) condition (44) is not satisfied for any $g(x)$.

Proof of Statement 1 is obtained by letting $v_o \to \infty$. It is, furthermore, obvious that Statement 2 holds if $p_n(v) > 0$ for all $v \leq v_o$.

Statements 1 and 2 indicate that if the amplitude probability density of the noise does not vanish on at least one-half of the real line, the average weighted error cannot be made zero except for a known message with constant amplitude. In Section 6.2 we discuss constraints on the message which rule out this trivial case. These constraints make it possible to find the messages that lead to a minimum nonzero error for a given noise. The problem of finding lower bounds on the error is thus meaningful for a large variety of noise characteristics.

The messages that lead to zero error when the noise is not of the type considered above do not offer a great interest either. We present two other statements for the sake of illustration.

STATEMENT 3: If $p_n(v)$ is such that

$$p_n(v) \begin{cases} > 0 & \text{for } v_o \leq v \leq v_1 \\ = 0 & \text{otherwise,} \end{cases}$$

then condition (52) is satisfied if and only if $p_m(v)$ consists of a set of impulses such that $p_n(x-v)$ for any x does not overlap two adjacent impulses. Hence, we have zero

53

error if and only if adjacent impulses in the message probability density are farther apart than $(v_1 - v_0)$. This statement is easily proved.

STATEMENT 4: If both message and noise have discrete probabilities, it is easy to verify that if

$$p_n(v) = \sum_k a_k u(v - d_k)$$

$$p_m(v) = \sum_j b_j u(v - \beta_j),$$

then condition (52) will be fulfilled if and only if

$$d_i - d_k \neq \beta_m - \beta_j \qquad \begin{cases} i \neq k \\ m \neq j. \end{cases}$$

The interpretation of these results is that we have to be able to assign one and only one value of the message to each region or point on the x axis of the input amplitudes to obtain zero error.

6.2 DISCUSSION OF CONSTRAINTS ON THE MESSAGE PROBABILITY DENSITY

Whenever the noise probability density does not vanish on at least one-half of the real line the average weighted error in a nonlinear no-memory filter can be made zero only if the message probability density is $p_m(v) = u(v - v_0)$, and, hence, the message has a known constant value. We wish to discuss here some constraints on the message probability density which have physical meaning and rule out such a trivial case. Our purpose is to define a set of conditions under which there is a lower bound on the error in filtering.

We consider first a message with a constant average power (say, equal to M^2). This constraint is clearly not sufficient here, and to rule out $p_m(v) = u(v - M)$, we have to require, also, that the mean value of the message be zero. This second condition will be satisfied if we consider messages of fixed average power which have an even probability density. Let us assume further that the noise probability density $p_n(v)$ is an even function of v and is nonzero for all values of v.

These conditions rule out the possibility of zero average weighted error. As we shall now show, however, by the use of a specific message satisfying the conditions stated, the average weighted error can still be made arbitrarily small, and hence no useful lower bound on the error can be obtained.

We consider the probability density of Fig. 19 for the message, and, furthermore, to satisfy the average-power constraint we require that $2aA^2 = M^2$.

We consider an error-weighting function $W(e)$ that is even and such that $W(0) = 0$. We shall show that, for most noises, the average weighted error $\overline{W(e)}$ can be made arbitrarily small for the specific message defined above. To show this we consider

54
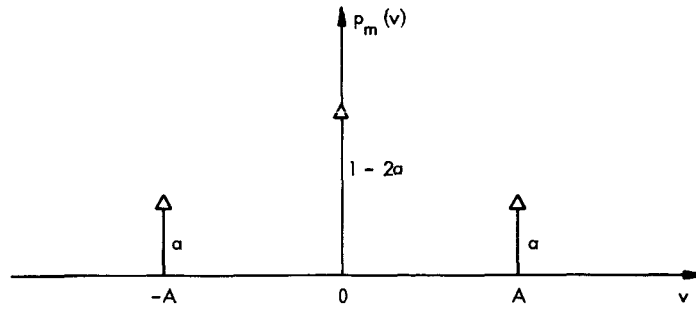
Fig. 19. Probability density considered.

the expression for the average weighted error for a nonoptimum filter $g(x)$ given by Eq. 49, which, since $p_{m+n/m}(x/v) = p_n(x-v)$, we write

$$\overline{W(e)} = \iint W[v-g(x)] \, p_m(v) \, p_n(x-v) \, dvdx. \tag{53}$$

For the message considered, we have

$$p_m(v) = (1-2a) \, u(v) + au(v-A) + au(v+A),$$

which leads to

$$\overline{W(e)} = a \int W[A-g(x)] \, p_n(x-A) \, dx + a \int W[-A-g(x)] \, p_n(x+A) \, dx$$

$$+ (1-2a) \int W[-g(x)] \, p_n(x) \, dx.$$

We now shall consider $\overline{W(e)}$ as $A \to \infty$ and the message power constraint is satisfied. If we can find some filter $g(x)$ such that $\overline{W(e)} \to 0$ as $A \to \infty$, then it is clear that if the optimum filter $g_{opt}(x)$ is used in each case, then $\overline{W(e)} \to 0$ as $A \to \infty$. This procedure, which avoids the use of an explicit expression for $g_{opt}(x)$, will yield necessary conditions on the behavior of $p_n(v)$ at $v \to \infty$, which are met for most error criteria and noises of interest.

Consider the nonoptimum filter characteristic

$$g(x) = \begin{cases} -A & x < -\dfrac{A}{2} \\[2mm] 0 & -\dfrac{A}{2} \leq x \leq \dfrac{A}{2} \\[2mm] A & x > \dfrac{A}{2}. \end{cases}$$

Then we have

$$W(e) = a \int_{-\infty}^{-A/2} W(2A) \, p_n(x-A) \, dx + a \int_{-A/2}^{A/2} W(A) \, p_n(x-A) \, dx$$

$$+ a \int_{A/2}^{\infty} W(0) \, p_n(x-A) \, dx + a \int_{-\infty}^{-A/2} W(0) \, p_n(x+A) \, dx$$

$$+ a \int_{-A/2}^{A/2} W(-A) \, p_n(x+A) \, dx + a \int_{A/2}^{\infty} W(-2A) \, p_n(x+A) \, dx$$

$$+ (1-2a) \int_{-\infty}^{-A/2} W(A) \, p_n(x) \, dx + (1-2a) \int_{-A/2}^{A/2} W(0) \, p_n(x) \, dx$$

$$+ (1-2a) \int_{A/2}^{\infty} W(-A) \, p_n(x) \, dx.$$

We now make use of the even character of $W(e)$ and $p_n(x)$ and of the property $W(0) = 0$ to write

$$\overline{W(e)} = 2a \int_{A/2}^{\infty} W(2A) \, p_n(x+A) \, dx + 2a \int_{-A/2}^{A/2} W(A) \, p_n(x+A) \, dx$$

$$+ 2(1-2a) \int_{A/2}^{\infty} W(A) \, p_n(x) \, dx.$$

By a simple change of variable in the integrals and by using $a = M^2/2A^2$, we write

$$\lim_{A \to \infty} \overline{W(e)} = \lim_{A \to \infty} \left[\frac{M^2}{A^2} \int_{A/2}^{3A/2} W(A) \, p_n(x) \, dx + \frac{M^2}{A^2} \int_{3A/2}^{\infty} W(2A) \, p_n(x) \, dx \right.$$

$$\left. + 2 \left(1 - \frac{M^2}{A^2} \right) \int_{A/2}^{\infty} W(A) \, p_n(x) \, dx \right].$$

By considering only the principle terms,

$$\lim_{A \to \infty} \overline{W(e)} = \lim_{A \to \infty} \left[\frac{M^2}{A^2} W(2A) \int_{3A/2}^{\infty} p_n(x) \, dx + 2W(A) \int_{A/2}^{\infty} p_n(x) \, dx \right].$$

Under most circumstances the limit on the right-hand side will go to zero. More precisely, if the noise has a finite variance, then we can use the Tchebycheff inequality in the form

$$\int_{B}^{\infty} p_n(x) \, dx \leq \frac{\sigma_n^2}{2B^2}$$

and we have

$$\lim_{A \to \infty} \overline{W(e)} \leq \lim_{A \to \infty} W(2A) \frac{2M^2 \sigma^2}{9A^4} + W(A) \frac{8\sigma_n^2}{A^2}.$$

If for any $\epsilon > 0$, the error-weighting function is such that $W(e) \leq \epsilon e^2$, then $\lim \overline{W(e)} \leq a$, where a denotes any arbitrarily small number.

For mean-square filtering ($W(e) = e^2$) it appears that a finite noise variance might not be sufficient to lead to an arbitrarily small error $\overline{W(e)}$ as $A \to \infty$. It is shown in Appendix D that a finite noise variance is in fact sufficient to give the result stated. This tighter result requires the use of the expression of the optimum filter $g_{opt}(x)$ for mean-square filtering.

More generally, if $p_n(x) \underset{x \to \infty}{\leq} \epsilon x^{-k}$, then it can be shown that $\lim_{A \to \infty} \overline{W(e)} \leq a$ if the error-weighting function is such that $|W(e)| \underset{e \to \infty}{\leq} \epsilon e^{k-1}$. If the noise behaves exponentially at infinity (e. g., Gaussian noise), then the error will be arbitrarily small for any error-weighting function of algebraic type at infinity.

We have just shown that, if we consider (i) a noise with a known, even probability density $p_n(v)$ which does not vanish for any value of v, (ii) a message with a known average power but an arbitrary even probability density, and (iii) an error-weighting function that is even and such that $W(0) = 0$, then no message can be found to satisfy the constraints which leads to zero average weighted error $\overline{W(e)}$, but $\overline{W(e)}$ can be made arbitrarily small by taking the message probability density shown in Fig. 19 and letting $A \to \infty$. This second fact was proved under some restriction on the error-weighting function $W(e)$ and the behavior on the noise probability density $p_n(v)$ for $v \to \infty$.

We now have to formulate an additional constraint on the message probability density which will lead to a nonzero lower bound on the error in filtering.

This additional constraint is suggested by considering the message probability density of Fig. 19 as $A \to \infty$. We note that, in the limit, we have a message of possibly infinite amplitude. This leads us to constrain the maximum amplitude of the message, as well as its average power.

VII. MAXIMUM AND MINIMUM OF THE AVERAGE WEIGHTED ERROR FOR A GIVEN NONLINEAR NO-MEMORY FILTER UNDER CONSTRAINTS

7.1 CONDITIONS UNDER WHICH FILTERING IS CONSIDERED

We consider filtering whenever the following conditions are met.

(i)   The amplitude probability density of the noise is a known even function.

(ii)   The characteristic of the nonlinear no-memory filter $g(x)$ is a known odd function.

(iii) The message has a constant average power M^2 and an amplitude less than a constant L.

(iv) The amplitude probability density of the message is even.

(v)   The error-weighting function $W(e)$ is even.

We wish to find the message probability density that maximizes or minimizes the average weighted error under these conditions.

7.2 MESSAGE PROBABILITY DENSITY MADE UP OF FOUR IMPULSES AT MOST

We show that the average weighted error will be minimized or maximized by a message amplitude probability density made up of four impulses at most. We consider the expression for the average weighted error when the message and the noise are statistically independent (Eq. 53). We can write Eq. 53 as

$$\overline{W(e)} = \int f(v)\, p_m(v)\, dv, \tag{54}$$

if we let

$$f(v) = \int W[v-g(x)]\, p_n(x-v)\, dx, \tag{55}$$

where $f(v)$ is the average weighted error corresponding to a specific value of v of the message amplitude and under our assumptions is a known function of v. Furthermore, it is simple to show, because of conditions (i), (ii), and (v), that $f(v)$ is an even function of v, and we can write

$$\overline{W(e)} = 2 \int_0^\infty f(v)\, p_m(v)\, dv.$$

Let us find the message probability density that minimizes $W(e)$ under power and amplitude constraints (as stated). With all conditions considered, our problem is to find the minimum (or maximum) of

$$\overline{W(e)} = \int_0^L f(v)\, 2p_m(v)\, dv$$

under the constraints

$$\int_0^L v^2 \, 2p_m(v) \, dv = M^2$$

$$\int_0^L 2p_m(v) \, dv = 1.$$

This problem is considered in Appendix E and use can be made of the results obtained there. We have, therefore, the fact that $\overline{W(e)}$ will be minimized (or maximized) by taking $2p_m(v)$ for $0 \leqslant v \leqslant L$ to be composed of at most two impulses. Since $p_m(v)$ is an even probability density, we have established the result that we have to consider a message probability density made up of four impulses at most.

7.3 DETERMINATION OF THE POSITION OF THE FOUR IMPULSES THAT LEAD TO A MAXIMUM OR A MINIMUM

We consider now an even message probability density made up of four impulses.

$$p_m(v) = \frac{a}{2} u(v-x) + \left(\frac{1}{2} - \frac{a}{2}\right) u(v-y) + \frac{a}{2} u(v+x) + \left(\frac{1}{2} - \frac{a}{2}\right) u(v+y), \tag{56}$$

in which

$$0 \leqslant a \leqslant 1 \tag{57}$$

$$x, y \leqslant L. \tag{58}$$

The power constraint becomes

$$ax^2 + (1-a) y^2 = M^2. \tag{59}$$

If, without loss of generality, we take x to be less than or equal to y, we have, necessarily,

$$0 \leqslant x \leqslant M$$
$$M \leqslant y \leqslant L. \tag{60}$$

We now need to determine x and y, the positions of the two impulses for $v \geqslant 0$, and a, the parameter for the magnitude of the impulses.

In the present situation the expression for the average weighted error (Eq. 54) becomes

$$\overline{W(e)} = a f(x) + (1-a) f(y), \tag{61}$$

which is to be minimized or maximized with respect to x, y, and a under constraints (57), (59), and (60).

a. General Solution

We use the method of Lagrange multipliers to find the extrema of

$$I = a\,f(x) + (1-a)\,f(y) + a\lambda x^2 + (1-a)\,\lambda y^2 - \lambda M^2,$$

in which λ is the Lagrange multiplier.

We set the three partial derivatives equal to zero.

$$\frac{\partial I}{\partial x} = 0 \rightarrow a[f'(x)+2\lambda x] = 0$$

$$\frac{\partial I}{\partial y} = 0 \rightarrow (1-a)[f'(y)+2\lambda y] = 0$$

$$\frac{\partial I}{\partial z} = 0 \rightarrow f(x) - f(y) + \lambda(x^2-y^2) = 0.$$

We see that whenever $a \neq 0$ and $a \neq 1$, these conditions are independent of a. Let us consider the cases for which $a = 0$ and $a = 1$ first.

For $a = 0$, we have $\frac{\partial I}{\partial x} = 0$ and we have to solve

$$f'(y) + 2\lambda y = 0$$

$$f(x) - f(y) + \lambda(x^2-y^2) = 0.$$

Let $y(\lambda)$, $x(\lambda)$ be solutions. Then the power constraint calls for $y(\lambda) = M$, and, therefore,

$$\lambda = -\frac{f'(M)}{2M},$$

and x has to be determined by the equation

$$f(x) - f(M) - \frac{f'(M)}{2M}(x^2-M^2) = 0.$$

Here, $x = M$ is clearly a solution; hence, for $a = 0$, $x = y = M$ we always have a stationary value of $\overline{W(e)}$.

For $a = 1$ it is clear, by symmetry, that $x = y = M$ corresponds also to a stationary value of $\overline{W(e)}$.

If $a \neq 0$, $a \neq 1$, our equations are independent of a and we have to solve formally the three equations for x and y as a function of λ (this is not necessarily possible), and then determine λ to satisfy the power constraint. But, since a is still available, we have the alternative method of solving the three equations for x, y, and λ. We know that for any x and y in the ranges specified there is always an a $(0 \leq a \leq 1)$ to satisfy the power constraint. We write the three equations

$$f'(x) = -2\lambda x \qquad 0 \leq x \leq M \tag{62}$$

$$f'(y) = -2\lambda y \qquad M \leq y \leq L \tag{63}$$

$$f(y) - f(x) = -\lambda(y^2-x^2). \tag{64}$$

Any λ such that x and y are in the indicated ranges is acceptable. Let us rewrite

Eq. 64 in the form

$$\int_x^y f'(v) \, dv = -2\lambda \int_x^y v \, dv. \tag{65}$$

A simple geometric interpretation can be given to Eqs. 62, 63, and 65 by considering a graph of $f'(v)$ (Fig. 20).

The points x $(0 \leqslant x \leqslant M)$ and y $(M \leqslant y \leqslant L)$ to be found are such that the two cross-hatched areas are equal. Since they are at the intersection of $f'(v)$ with the line -2λ, it is clear that Eqs. 62, 63, and 65 are then satisfied, and x and y correspond to a stationary value of $\overline{W(e)}$.

Another interpretation of Eqs. 62-64 is to note that they require $f(v)$ to be tangent to the parabola $h(v) = -\lambda v^2 + C$ at points x and y corresponding to a stationary value of $\overline{W(e)}$. Here, C denotes an arbitrary constant. This interpretation is, in fact, more fruitful in the discussion of extrema of $W(e)$, as well as stationary values. To see this, consider the expression to be minimized (Eq. 61); when, in fact,

$$f(v) = -\lambda v^2 + C,$$

we obtain

$$\overline{W(e)} = a[-\lambda x^2 + C] + (1-a)[-\lambda y^2 + C] = -\lambda[ax^2 + (1-a)y^2] + C$$

and, since the constraint requires that $ax^2 + (1-a)y^2 = M^2$, we have

$$\overline{W(e)} = -\lambda M^2 + C.$$

Hence, the parabola $h(v) = -\lambda v^2 + C$ is the curve corresponding to a constant value of $W(e)$, the expression to be minimized. In other words, if the positions of the two impulses are x and y, and if we find λ and C such that $f(x) = -\lambda x^2 + C$ and $f(y) = -\lambda y^2 + C$, then $\overline{W(e)} = -\lambda M^2 + C$. It is important to note that $\overline{W(e)}$ is then the ordinate of the parabola $h(v) = -\lambda v^2 + C$ at point $v = M$. Therefore, the problem of minimization (or maximization) considered can be discussed in the following terms.
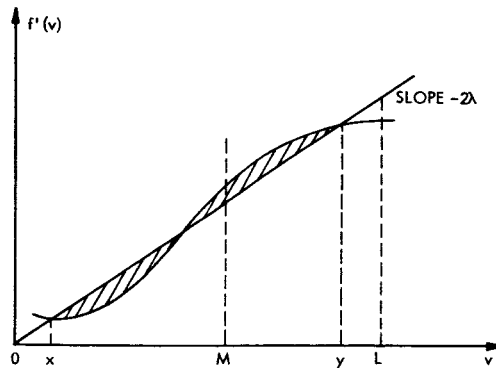

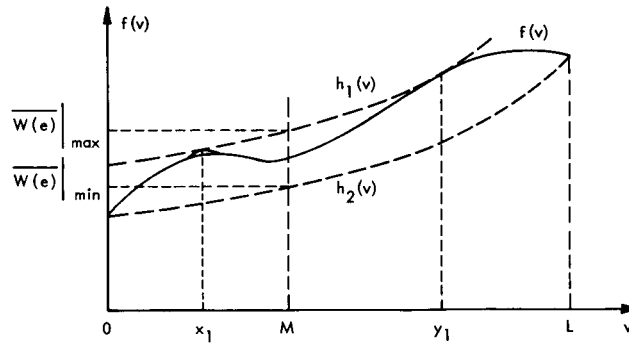Fig. 20. Geometric interpretation of Eqs. 62, 63, and 65.

Fig. 21. Other interpretation of Eqs. 62-64.

Given the curve f(v), which corresponds to the average weighted error when the message has value v, find the parabola $h(v) = -\lambda v^2 + C$ intersecting f(v) between 0 and M and between M and L such that $\overline{W(e)} = -\lambda M^2 + C$ is minimum (or maximum). The intersecting points x and y give, then, the positions of the impulses making up the message probability density. The situation is illustrated in Fig. 21.

The parabola corresponding to the maximum average weighted error $\overline{W(e)}_{max}$ is $h_1(v)$. It is tangent to f(v) at the two points of abscissae x_1 and y_1. The minimum is given by $h_2(v)$, which intersects f(v) at points x = 0 and y = L. Our previous analytical results now become clear.

(i) If the parabola $h(v) = -\lambda v^2 + C$ is tangent to f(v) at points x and y, then we have there a stationary value of $\overline{W(e)}$.

(ii) The points x = y = M always give a stationary value to $\overline{W(e)}$ when we can draw a parabola tangent in v = M to f(v).

b. Cases for Which the Extrema Correspond to Points on the Boundaries

The discussion of these cases becomes easy if we make use of a property of the family of parabolas considered. We shall state this property in the form of a lemma.

LEMMA: Consider the family of parabolas $z(v) = -\lambda v^2 + C$ and a specific parabola $z_1(v)$ passing through the point (v = M, $z_1(M) = -\lambda M^2 + C = G$), in which G is a given constant. Then, if $z(v) \neq z_1(v)$, we have either

$$z(v) < z_1(v) \qquad 0 \leq v \leq M$$

or

$$z(v) < z_1(v) \qquad v \geq M.$$

PROOF: Two parabolas of the family intersect, at most, at one point for $v \geq 0$. Hence, if $z_1(M) = G$ and $z(M) \leq G$, then z(v) and $z_1(v)$ intersect once at most; either $0 \leq v \leq M$ or $v \geq M$; hence, z(v) is necessarily below $z_1(v)$ in one of the two regions.

Q. E. D.

There are four cases for which the minimum of W(e) occurs on a boundary.

CASE 3

$x = 0$, $y = L$ corresponds to the minimum value of $\overline{W(e)}$ if the parabola $h_1(v) = \dfrac{f(L) - f(0)}{L^2} v^2 + f(0)$, which has the properties that $h_1(0) = f(0)$ and $h_1(L) = f(L)$, is such that $f(v) \geq h_1(v)$ for $0 \leq v \leq L$. The minimum average weighted error is then

$$\overline{W(e)}_{min} = \left[f(L) - f(0) \right] \frac{M^2}{L^2} + f(0). \tag{66}$$

CASE 4

$x = y = M$ corresponds to the minimum value of $\overline{W(e)}$ if the parabola $h_2(v) = \dfrac{f'(M)}{2M} (v^2 - M^2) + f(M)$, tangent to $f(v)$ at $v = M$, is such that $f(v) \geq h_2(v)$ for $0 \leq v \leq L$. The minimum average weighted error then is

$$\overline{W(e)}_{min} = f(M).$$

These cases can be proved as follows. There is no other pair of points x and y satisfying the constraints which would lead to a smaller average weighted error. Such a pair of points would have to be on a parabola of the form $h(v) = -\lambda v^2 + C$ such that $h(M) < h_1(M)$ or $h(M) < h_2(M)$. From the lemma all such parabolas will not intersect $f(v)$, either for $0 \leq v \leq M$ or for $M \leq v \leq L$.

CASE 5

$x = 0 \qquad M < y < L$

CASE 6

$0 < x < M \qquad y = L$

Cases 5 and 6 can be similarly discussed without difficulty.

The discussion of the position of the impulses leading to a maximum of the average weighted error is completely similar. One has, then, to consider parabolas $h(v)$, which intersect $f(v)$ twice, for $0 \leq v \leq L$ and $M \leq v \leq L$, and are such that $h(v) \geq f(v)$, $0 \leq v \leq L$.

It is often clear, by considering the graph of $f'(v)$, that the minimum will occur at $x = y = M$ and the maximum at $x = 0$, $y = L$, or conversely.

For instance, if $f'(v)$ and $-2\lambda v$ (any λ) intersect only once for $0 \leq v \leq L$, it is easy to see that a parabola $h(v) = -\lambda v^2 + C$ tangent at M cannot intersect $f(v)$ for $0 \leq v \leq L$, and if a parabola $h(v)$ intersects at $x = 0$, $y = L$, then it will not intersect anywhere else in the range. Note that if $x = 0$, $y = L$ corresponds to an extremum of $\overline{W(e)}$ for some M ($0 \leq M \leq L$), then x and y will keep this property for any M in that range.

7.4 AVERAGE WEIGHTED ERROR FOR A GIVEN MESSAGE AMPLITUDE AT THE INPUT

We establish three properties of the functions $f(v)$, the average weighted error when the message amplitude is v, since this function plays a central role in our previous discussion.

a. f(v) and Its Derivative at the Origin

From Eq. 55, we compute $d\,f(v)/dv$ by interchanging the order of differentiation and integration; we have

$$\frac{df(v)}{dv} = \int W'[v-g(x)]\, p_n(x-v)\, dx - \int W[v-g(x)]\, p'(x-v)\, dx,$$

in which the prime indicates differentiation with respect to the argument. Hence,

$$\left.\frac{df(v)}{dv}\right|_{v=0} = \int W'[-g(x)]\, p_n(x)\, dx - \int W[-g(x)]\, p'(x)\, dx.$$

If $W(e)$ is even, then $W'(e)$ is odd. Then, since $g(x)$ is odd, $W(-g(x))$ is even and $W'(-g(x))$ is odd. And, since $p_n(x)$ is even, $p'_n(x)$ is odd. We see, therefore, that each of the integrals is zero, and we have

$$\left.\frac{df(v)}{dv}\right|_{v=0} = 0.$$

We have stated that, if the line $-2\lambda v$ crosses $f'(v)$ only once in the range $0 \leqslant v \leqslant L$, then the pairs $(x = 0,\ y = L)$ and $(x = y = M)$ correspond to extrema of the average weighted error $\overline{W(e)}$. Now, since $f'(0) = 0$, we have a crossing point at $v = 0$. If $f'(v)$ is continuous with continuous derivatives, then $f'(v)$, either concave or convex, will guarantee that we have only one crossing with a straight line going through the origin, besides the point $v = 0$. Hence, we have now the weaker but simpler sufficient condition that, if either

$$\frac{d^3 f(v)}{dv^3} \geqslant 0 \qquad 0 \leqslant v \leqslant L$$

or

$$\frac{d^3 f(v)}{dv^3} \leqslant 0 \qquad 0 \leqslant v \leqslant L$$

then the extrema of $\overline{W(e)}$ will occur when x and y are at the extrema of their ranges.

b. The Function f(v) When W(e) Is a Polynomial in e and g(x) Is a Polynomial in x

If $W(e)$ is a polynomial of the error and the given filter $g(x)$ is a polynomial in x, then $f(v)$ is obtained simply in terms of the moments of the noise probability density.
    Let

$$W(e) = \sum_{k=0}^{2n} a_k\, e^k$$

$$g(x) = \sum_{j=1}^{r} b_j x^j$$

and write, by a change of variable,

$$f(v) = \int W[v-g(x+v)] \, p_n(x) \, dx \tag{67}$$

and we have

$$W[v-g(x+v)] = \sum_k a_k \left[v - \sum_j b_j(x+v)^j \right]^k . \tag{68}$$

By expanding all terms, Eq. 68 can be written

$$W[v-g(x+v)] = \sum_q \sum_\ell C_{\ell q} \, v^\ell x^q .$$

Since we average this expression with respect to x, we have

$$f(v) = \sum_q \sum_\ell C_{\ell q} \, M_q \, v^\ell ,$$

in which M_q is the moment of $p_n(x)$ of order q,

$$M_q = \int x^q p_n(x) \, dx .$$

Since $p_n(x)$ is even, all odd moments are zero. If 2n is the degree of the error polynomial $W(e)$ and r is the degree of the given filter $g(x)$, then $f(v)$ will be a polynomial of degree 2nr, involving moments of $p_n(x)$ up to order 2nr.

c. Expression of f(v) for Criterion d

Using criterion d and Eq. 67, we have

$$W[v-g(x+v)] = \begin{cases} 0 & |v-g(x+v)| < A \\ 1 & |v-g(x+v)| \geq A. \end{cases}$$

The condition

$$|v-g(x+v)| < A$$

can be written

$$v - A < g(x+v) < v + A.$$

Assume that g(x) has an inverse $g^{-1}(x)$; then the inequalities take the form

$$g^{-1}(v-A) < x + v < g^{-1}(v+A).$$

From the way in which the inequalities are written we assume that $g(x)$ is an increasing function of x. If it is not, the two limits have to be interchanged.

Hence, in terms of x we have

$$g^{-1}(v-A) - v < x < g^{-1}(v+A) - v,$$

and for this range of x, $W[v-g(x+v)] = 0$. Now $f(v)$ takes a simple form

$$f(v) = \int_{-\infty}^{g^{-1}(v-A)-v} p_n(x)\ dx + \int_{g^{-1}(v+A)-v}^{\infty} p_n(x)\ dx$$

or

$$f(v) = 1 - \int_{g^{-1}(v-A)-v}^{g^{-1}(v+A)-v} p_n(x)\ dx. \tag{69}$$

If $g(x)$ does not have an inverse, we shall have to consider several ranges of x for each value of v.

7.5 EXAMPLES OF THE DETERMINATION OF EXTREMA OF THE AVERAGE WEIGHTED ERROR

EXAMPLE 1: We consider a simple example employing the mean-square error criterion and Gaussian noise. Let

$$g(x) = x^3 \qquad W(e) = e^2 \qquad p_n(x) = \frac{1}{\sqrt{2\pi}\ \sigma} \exp\left[-\frac{x^2}{2\sigma^2}\right];$$

then

$$f(v) = \int [v-x^3]^2\ p_n(x-v)\ dx.$$

If we expand and average and use the expressions of the higher moments of the Gaussian probability density in terms of σ^2, we obtain

$$f(v) = v^6 + v^4(15\sigma^2-2) + v^2(45\sigma^4-6\sigma^2+1) + 15\sigma^6.$$

If we take $\sigma = 1$,

$$f(v) = v^6 + 13v^4 + 40v^2 + 15$$

and

$$f'(v) = 6v^5 + 52v^3 + 80v.$$

It is clear that $f'(v)$ is a convex function of v for all v. This fact indicates that $x = y = M$ is the position of the impulse in the message probability density leading to the minimum
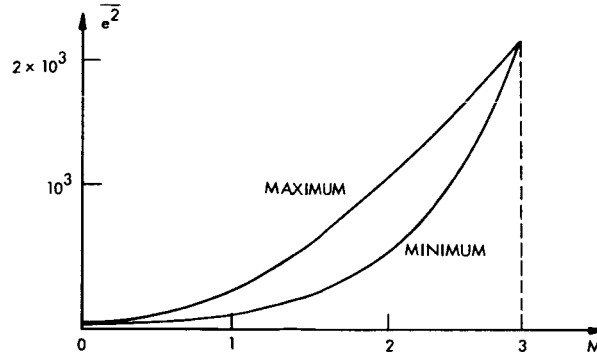
66

Fig. 22. Mean-square error for $g(x) = x^3$ and Gaussian noise.

mean-square error and that $x = 0$, $y = L$ corresponds to the maximum mean-square error. Therefore, we have

$$\overline{e^2_{min}} = f(M) = M^6 + 13M^4 + 40M^2 + 15$$

$$\overline{e^2_{max}} = [f(L)-f(0)] \, M^2/L^2 + f(0)$$

$$\overline{e^2_{max}} = (L^4 + 13L^2 + 40) \, M^2 + 15.$$

We give in Fig. 22 a graph of the numerical values $L = 3$ and $0 \leqslant M \leqslant 3$.

EXAMPLE 2: As an illustration of a non mean-square error criterion, we consider

$$g(x) = ax$$

$$p_n(x) = \frac{1}{2} e^{-|x|}$$

and criterion d, and undertake to find lower and upper bounds on the average error under the general conditions of section 7.1. Making use of Eq. 69, we have

$$f(v) = 1 - \frac{1}{2} \int_{\frac{v-A}{a} - v}^{\frac{v+A}{a} - v} e^{-|x|} \, dx,$$

where $f(v)$ is an even function. For $v \geqslant 0$ we have

$$f(v) = \begin{cases} e^{-A/a} \cosh \left(\frac{1}{a} - 1\right) v & 0 \leqslant v \leqslant \frac{A}{1 - a} \\[4mm] 1 - \sinh \frac{A}{a} \exp\left[-\left(\frac{1}{a} - 1\right)v\right] & v \geqslant \frac{A}{1 - a} \end{cases}$$

Let us take a, the attenuation constant of the filter, as a parameter, fix the values of the other parameters as

67

M = 2,   L = 4,   A = 0.2,

and consider the graph of f'(v) given in Fig. 23. In Fig. 23 the portion of the curve f'(v) for $0 \leqslant v \leqslant 0.2/(1-a)$ is convex and, therefore, if $L \leqslant 0.2/(1-a)$ we shall have x = y = M for the minimum value of $\overline{W(e)}$ and (x = 0, y = L) for the maximum value. Since f'(v) for $0 \leqslant v \leqslant 0.2/(1-a)$ is very close to its tangent at v = 0, abscissa L_o (defined in Fig. 23)
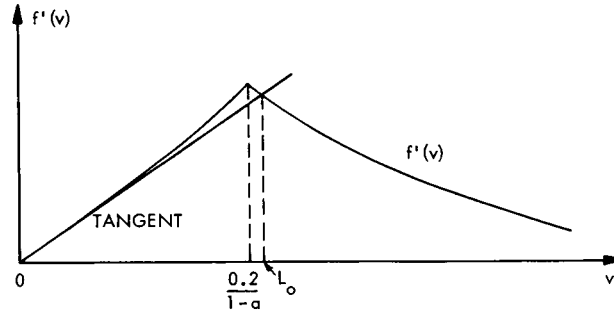
Fig. 23. Geometric discussion of the position of the impulses.

will be approximately equal to 0.2/(1-a). Under these conditions it is easy to verify that whenever $L > 0.2/(1-a)$ the pair (x = 0, y = L) will give $\overline{W(e)}_{min}$, and if $M > 0.2/(1-a)$ the pair (x = y = M) will give $\overline{W(e)}_{max}$. With the same approximation, when $M < 0.2/(1-a)$ and $L > 0.2/(1-a)$ the pair (x = 0, y = 0.2/(1-a)) will give $\overline{W(e)}_{max}$. These results are

Table 2. Position of the impulses for which $\overline{W(e)}$ is an extremum.

M \ L	0	$\dfrac{0.2}{1-a}$	∞
0			
	$\text{Min} \begin{cases} x = M \\ y = M \end{cases}$	$\text{Min} \begin{cases} x = 0 \\ y = L \end{cases}$	
	$\text{Max} \begin{cases} x = 0 \\ y = L \end{cases}$	$\text{Max} \begin{cases} x = 0 \\ y = \dfrac{0.2}{1-a} \end{cases}$	
$\dfrac{0.2}{1-a}$			
		$\text{Min} \begin{cases} x = 0 \\ y = L \end{cases}$	
∞		$\text{Max} \begin{cases} x = M \\ y = M \end{cases}$	

68

Fig. 24. Average weighted error for Example 2.
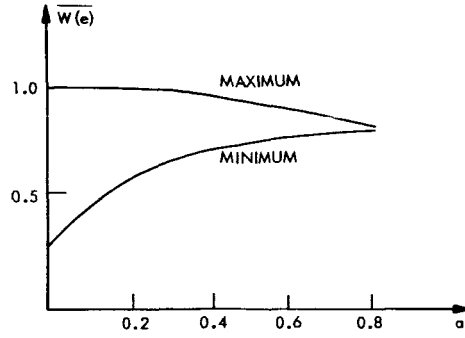
summarized in Table 2. The ranges for L and M defined in Table 2 correspond to ranges in a. Since M = 2 and L = 4, we have

$$L < \frac{0.2}{1-a} \to a > 0.95$$

$$M < \frac{0.2}{1-a} \to a > 0.9.$$

By the use of the expressions of $\overline{W(e)}_{max}$ and $\overline{W(e)}_{min}$ obtained previously we can write

$$\overline{W(e)}_{min} = \begin{cases} 0.25\left[1 - \sinh\frac{0.2}{a}\exp\left[-4\left(\frac{1}{a}-1\right)\right]\right] + 0.75\,e^{-0.2/a} & 0 \le a \le 0.95 \\ e^{-0.2/a}\cosh 2\left(\frac{1}{a}-1\right) & 0.95 \le a < 1 \end{cases}$$

$$\overline{W(e)}_{max} = \begin{cases} 1 - \sinh\frac{0.2}{a}\exp\left[-2\left(\frac{1}{a}-1\right)\right] & 0 \le a < 0.9 \\ 20(1-a)\left(\cosh\frac{0.2}{a}-1\right)e^{-0.2/a} + e^{-0.2/a} & 0.9 \le a < 0.95 \\ 0.25\,e^{-0.2/a}\left[\cosh\frac{4(1-a)}{a}-1\right] + e^{-0.2/a} & 0.95 \le a < 1 \end{cases}$$

In Fig. 24 we give a graph of these results for $0 \le a \le 0.8$.

7.6 EXTENSIONS OF THE RESULTS

a. Removal of the Maximum-Amplitude Constraint

In some cases the upper and lower bounds on the average weighted error will stay finite and nonzero, respectively, when L, the maximum amplitude of the message, goes to infinity. This will depend on the behavior of f(v), the average weighted error when the message has an amplitude v.

(i) Lower bound

We note that if f(v) > 0 for all v, then the lower bound on the average weighted error will be larger than zero for any L. We can say, more exactly, that $\overline{W(e)}_{min}$ will not be

69

smaller than f_{min}, the minimum value of $f(v)$ over all v, since $\overline{W(e)} = a f(x) + (1-a)f(x) \geq f_{min}$. Therefore, if $f(v) > 0$ for all v we are sure of finding some useful lower bound to the average weighted error for any M and any L.

(ii) Upper bound

If $f(v)$ is bounded for all finite v, the behavior of $f(v)$ for $v \to \infty$ will determine whether or not some finite upper bound on the average weighted error will exist as $L \to \infty$. Since we are looking for a parabola that is tangent to $f(v)$ from above, the problem is to determine whether or not we can find some parabola $h(v) = -\lambda v^2 + C$, with λ and C finite, such that $f(v) < h(v)$ for all v. It is clear that if $f(v) < k_1 v^2$ for some k_1 finite, then such a parabola $h(v)$, with λ and C finite, can be found. If $f(v) > k_2 v^2$ for any k_2, then no finite upper bound will exist. For illustration we consider Examples 1 and 2 and look at the average weighted error as $L \to \infty$. From the results of Example 1, $\overline{W(e)}_{min}$ is equal to $f(M)$ and does not depend on the maximum-amplitude constraint L. For the upper bound we have $\overline{W(e)}_{max} \to \infty$ as $L \to \infty$, since $f(v) \sim v^6$ as $v \to \infty$.

From Example 2 (Table 2) for $L \to \infty$ we have the fact that $\overline{W(e)}_{min}$ will be obtained by taking $x = 0$, $y \to \infty$. From Eq. 66 we see that if $M^2/L^2 \to 0$ as $L \to \infty$, then $\overline{W(e)}_{min} \to f(0) = e^{-0.2/a}$. By taking

$$\left. \begin{array}{l} x = 0 \\ \\ y = \dfrac{0.2}{1-a} \end{array} \right\} \quad \text{if } M \leq \dfrac{0.2}{1-a}$$

$$x = y = M \qquad \text{if } M > \dfrac{0.2}{1-a},$$

we obtain

$$\overline{W(e)}_{max} = \begin{cases} \left[f\!\left(\dfrac{0.2}{1-a}\right) - f(0) \right] \dfrac{M^2}{\left(\dfrac{0.2}{1-a}\right)^2} + f(0) & M \leq \dfrac{0.2}{1-a} \\ \\ f(M) & M > \dfrac{0.2}{1-a}. \end{cases}$$

b. Other Constraints of the Form $\int F(x)\, p_m(x)\, dx = F$

We have been concerned thus far in this section with an average power constraint on the message. The results obtained can be easily extended to other constraints of the form $\int F(x)\, p_m(x)\, dx = F$. The key result is that for all constraints of this type we need only consider a message probability density made up of four impulses. This allows us to duplicate closely our previous reasoning. Since we left unchanged all other conditions stated in section 7.1, we are still looking for the positions of x and y of two impulses such that

$$\overline{W(e)} = a f(x) + (1-a) f(y)$$

70

is maximum or minimum under the constraint

$$a F(x) + (1-a) F(y) = F,$$

and we require that $0 \leq a \leq 1$ and $0 \leq x, y \leq L$.

The curves $h(v) = \lambda F(v) + C$ correspond to constant values of $\overline{W(e)}$. That is, if x and y are chosen to give $f(x)$ and $f(y)$, and if we select λ and C such that $h(x) = f(x)$, $h(y) = f(y)$, then the corresponding average weighted error is $\overline{W(e)} = \lambda F + C$.

This is easy to see, since if $f(v) = h(v) = \lambda F(v) + C$, then, for all x and y satisfying the constraint, we have $\overline{W(e)} = \lambda F + C$.

Now, we wish to prove that if $F(v)$ is a continuous, monotonically increasing or decreasing function of v, then, for a constraint of the form $a F(x) + (1-a) F(y) = F$ with $0 \leq x, y \leq L$, and $0 \leq a \leq 1$, there is a value v_o such that $F(v_o) = F$. Furthermore, for any x and y, $0 \leq x \leq v_o$ and $v_o \leq y \leq L$, the constraint can always be satisfied by proper choice of a.

PROOF: Since $0 \leq a \leq 1$, we have either $F(x) \leq F$ and $F(y) > F$ or the converse; and, since $F(v)$ is continuous, there is a v_o, $x < v_o < y$ such that $F(v_o) = F$. Now take $x < v_o$, $y > v_o$; then $F(x) < F$ and $F(y) > F$ or the converse. Therefore, since $a = [F-F(y)]/[F(x)-F(y)]$, we have $0 \leq a \leq 1$.

Therefore, the analogy with the average power constraint is complete. We have to find a curve of the family $h(v) = \lambda F(v) + C$ which intersects $f(v)$ in two points x and y with $0 \leq x \leq v_o$ and $v_o \leq y \leq L$ and has the smallest ordinate $h(v_o) = \lambda F(v_o) + C$, and, furthermore, we have $F(v_o) = F$.

EXAMPLE: We consider the constraint $\int |x| \, p_m(x) \, dx = F$. The family of curves $h(v)$ is here $h(v) = \lambda v + C$ for $v \geq 0$, and we have $v_o = F$. Therefore we wish to find a straight line $h(v) = \lambda v + C$ that intersects $f(v)$ in two points x and y ($0 \leq x \leq F$ and $F \leq y \leq L$) and is such that $h(F) = \lambda F + C$ is maximum or minimum.

Stationary values of $\overline{W(e)}$ will occur if a line is tangent at two points to $f(v)$. If $f(v)$ is convex, then $x = y = F$ will give the minimum average weighted error and $x = 0$, $y = L$ will give the maximum, and if $f(v)$ is concave the converse will be true.

c. Removal of the Requirements of an Odd Filter Characteristic and an Even Message Probability Density

We found that an even noise probability density, an odd filter characteristic, and an even error-weighting function resulted in the function $f(v)$ being even, which is the property of $f(v)$ that we needed in our discussion. If we do not restrict $f(v)$ to be even, then none of the restrictions discussed above on noise, filter, and error-weighting function are needed. In terms of $f(v)$ we can rewrite the conditions of section 7.1 as follows.

(i') The function $f(v)$ is even.

(ii') The message has a constant average power, M^2, and an amplitude less than a constant L.

(iii') The amplitude probability density of the message is even.

We wish to remove here conditions (i') and (iii'), leaving unchanged condition (ii') just discussed.

The results are modified quite simply to remove the even requirement on the message probability density. In terms of the parabolas that determine the position of the four impulses it is clear, because of the even character of conditions (i') and (ii'), that the position of the four impulses stays even. The impulses at $-x$ and $+x$, however, are no longer required to have the same area. In fact, only the sum of their areas is important, for if $f(v)$ is even, then

$$a_1 f(x) + a_2 f(-x) + a_3 f(y) + a_4 f(-y) = (a_1 + a_2) f(x) + (a_3 + a_4) f(y).$$

It is clear that an even message probability density is among the solutions and that all solutions lead to the same average weighted error. If the message and the noise can take only positive values, we can also immediately use our previous results. In this new situation the function $f(v)$ is only of interest for $v \geq 0$, and, by assuming that it is even, we can discuss the problem as before. In the results we give zero area to the impulses corresponding to a negative message amplitude.

If we do not require $f(v)$ or $p_m(v)$ to be even any more, then, from Appendix E, the extrema of the average weighted error can still be found for a message probability density made up of two impulses at most. If x and y are the positions of the impulses, we still require that $0 \leq |x| \leq M$, $M \leq |y| \leq L$, and M, $L > 0$, because of the power constraint. The stationary values of $\overline{W(e)}$ still correspond to a parabola $h(v) = -\lambda v^2 + C$ tangent to $f(v)$ in x and y, and tangency is determined by considering a graph of $f'(v)$ and its intersections with the line $h'(v) = -2\lambda v$. Here, again, the extension is quite straightforward.

EXAMPLE: We take

$$g(x) = x^2 \qquad W(e) = e^2 \qquad p_n(v) = e^{-v^2}/\sqrt{\pi},$$

which lead to

$$f(v) = \int (v-x^2)^2 \, p_n(x-v) \, dx$$

$$f(v) = v^4 - 2v^3 + 4v^2 - v + 3/4$$

and

$$f'(v) = 4v^3 - 6v^2 + 8v - 1.$$

It is simple to show that $f'(v)$ is monotonically increasing and that a line $h'(v) = -2\lambda v$ will only intersect $f'(v)$ at one point. This means that a parabola $h(v) = -\lambda v^2 + C$ cannot be tangent to $f(v)$ at more than one point. Furthermore, it can be shown that such a parabola will be below $f(v)$. The lower bound will be found by taking $x = y = +M$ or $x = y = -M$. Similarly, for the upper bound we take $x = 0$, $y = +L$ or $x = 0$, $y = -L$. Since $f(-v) > f(v)$, the proper choice is $x = y = +M$ for the lower bound and $x = 0$, $y = -L$ for the upper bound, and the lower bound is given by $f(M)$ independently of L.

VIII. LOWER BOUNDS ON THE AVERAGE WEIGHTED ERROR
IN A NONLINEAR NO-MEMORY FILTER

8.1 INTRODUCTION

We saw in Section III that, whenever the message and noise probability densities are known, it is possible to determine the nonlinear no-memory filter that is optimum for a specific error criterion. In Section VI we discussed conditions under which the average weighted error cannot be made zero and useful lower bounds exist. By combining this information with the results of Section VII, we shall give here lower bounds to the average weighted error for specific error criteria. We consider filtering under the following conditions.

(i)   We only know of the message that its probability density $p_m(v)$ is even, with a known variance M^2 and $p_m(v) = 0$ for all $|v| > L$.

(ii) The probability density of the noise $p_n(v)$ is a known even function, nonzero for all v and nonincreasing for $v \geq 0$.

(iii) The error-weighting function $W(e)$ is a known, nondecreasing function of e for $e \geq 0$ and even.

The first problem is to find the specific message having the given characteristics which leads to a minimum average weighted error. For this message we then consider the average weighted error for various error criteria.

8.2 OPTIMUM MESSAGE PROBABILITY DENSITY WHEN $p_n(v)$ AND $W(e)$ ARE EVEN

We are looking for the specific message probability density that meets condition (i) and leads to the minimum average weighted error. Conditions (ii) and (iii) can be relaxed, and we require only that $p_n(v)$ and $W(e)$ be known and even. This problem differs from that of Section VII in that the filter is not known and we now are minimizing among the filters, as well as among the message probability densities. We have shown that if the filter is known and odd the minimum average weighted error is obtained when the message probability density is made up of 4 impulses at most. We shall show here that if minimization is carried out among the filters as well we are still led to consider a message probability density made up of 4 impulses at most. To show this we need the result, which follows from symmetry considerations, that when the message probability density, the noise probability density, and the error-weighting function are even, then the optimum nonlinear no-memory filter is odd.

Let us assume that we know $p_1(v)$, the message probability density giving the minimum average weighted error, and $g_1(x)$, the corresponding optimum filter. Here, again, we can write the average weighted error in the form

$$\overline{W_1(e)} = \int f_1(v) \, p_1(v) \, dv,$$

in which

$$f_1(v) = \int W[v-g_1(x)] \, p_n(x-v) \, dx.$$

Since $g_1(x)$ and therefore $f_1(v)$ are known, we can minimize among the message probability densities, keeping $f_1(v)$ fixed. From Section VII we can find here a message probability density $p_2(v)$ made up of 4 impulses at most, and we obtain $\overline{W_2(e)}$ as the resulting average weighted error.

We now can start from $p_2(v)$, consisting of 4 impulses at most, and find the optimum filter $g_2(x)$ and the minimum average weighted error $\overline{W_3(e)}$. From the way in which $\overline{W_1(e)}$, $\overline{W_2(e)}$, and $\overline{W_3(e)}$ have been successively obtained, it is clear that we have

$$\overline{W_3(e)} \leqslant \overline{W_2(e)} \leqslant \overline{W_1(e)}. \tag{70}$$

Since, by hypothesis, $\overline{W_1(e)}$ is the lowest achievable average weighted error for this class of messages, it is necessary that we have equality in Eq. 70. But $\overline{W_3(e)}$ corresponds to a message probability density made up of 4 impulses at most and, therefore, the minimum of $\overline{W(e)}$ will be found by considering only messages consisting of 4 impulses at most.

8.3 CONDITIONS FOR WHICH THE OPTIMUM MESSAGE PROBABILITY DENSITY CONSISTS OF THREE IMPULSES AT MOST

In this section we justify, on intuitive grounds, a result that we are not able to prove rigorously. Since our argument is intuitive, it is helpful before discussing the specific point to illustrate more physically the result of Section VII.

We consider, for instance, any message that satisfies the conditions

(i) The message changes at discrete times $t = 1, 2, \ldots$ and then takes any amplitude less than L.

(ii) The message has an even amplitude probability density and a known average power.

Let the additive noise be white Gaussian and assume that delay in filtering is
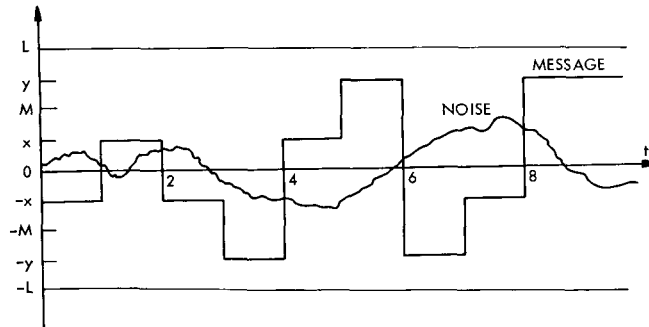

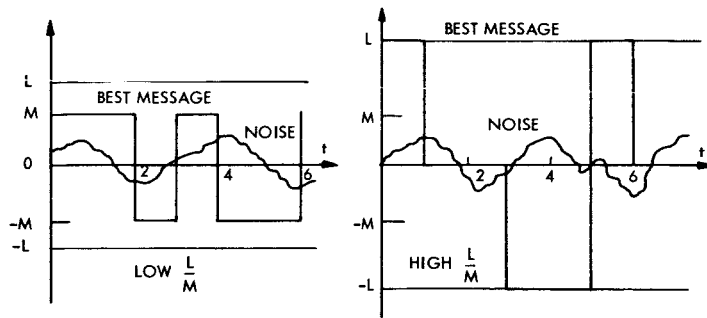Fig. 25. Optimum message for additive white Gaussian noise.

74

Fig. 26. Message levels for which the distance between
levels is maximum.

acceptable. Then the best filtering operation is to integrate the input from time $t = n$, $n = 1, 2, \ldots$ to time $t = n + 1$ (matching the filter to the message) and then perform no-memory filtering to estimate the amplitude of the message in that time interval. We saw in section 8.2 that the message that is most distinguishable from the noise, in the mean-square sense, for instance, is quantized at no more than 4 different levels. This situation is illustrated in Fig. 25.

We shall discuss this result in physical terms: We showed in Section VI the quite intuitive result that to obtain zero error we have to be able to assign a unique value of the message to each value of the input. In general, the effect of the noise will be to smudge the amplitude probability of the message and make impossible a one-to-one correspondence between input amplitudes and message amplitudes. When we allow some choice among the message probability densities and we wish to reduce the confusion between the possible messages which is due to the noise, it is clear that a small number of amplitude levels for the message will be beneficial. It is, in fact, the result of section 8.2 that, by proper choice among a maximum of 4 message amplitudes, we obtain the minimum average weighted error that is possible for messages of the class considered.

Another intuitive notion is that, for quite general conditions, what should be done is to increase the distance between the possible messages as much as the message constraints will allow. For the constraints on the average power and the peak power that we consider here (section 8.1), the maximum distance between messages will be given by a message quantized at either 3 or 2 different levels, as illustrated in Fig. 26.

Some restrictive conditions on $p_n(v)$ and $W(e)$ are needed, and we shall indicate why conditions (ii) and (iii) of section 8.1 appear to be necessary. First, note that conditions (ii) and (iii) allow most cases of interest. By considering specific examples, we show that conditions (ii) and (iii) are needed to rule out the cases in which some specific positions of the 4 impulses making up the message probability density are highly favored, because of the error-weighting function selected or because of peculiarities of the noise probability density. To illustrate this specifically, let us assume, for the quantized message considered, that the filter will give an output quantized at the same levels. This
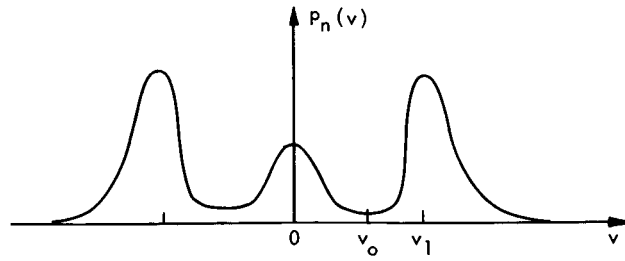
75

Fig. 27. Noise probability density for which condition (ii) is not fulfilled.

assumption simplifies our discussion, and it does not invalidate our argument to con-
sider a filter possibly nonoptimum. The average weighted error will depend on two
effects: (a) We have a certain probability, tied to the noise characteristic, of making
an error, i. e., to select the wrong message. (b) We give a certain weight, determined
by W(e), to the magnitude of this error.

Assume, first, that condition (ii) is not fulfilled; $p_n(v)$ is even but may increase for
$v \geq 0$. A possible case is illustrated in Fig. 27. If one of the message amplitudes is x,
it is quite probable that the input amplitude will be approximately $x \pm v_1$ and improbable
that it will be approximately $x \pm v_0$. This will favor $x \pm v_0 \pm v_1$ as the positions of the
other message impulses, in order to lead to the desired goal that for each region of the
input only one value of the message be most probable. If, as required by condition (ii)
$p_n(v)$ does not increase for $v \geq 0$, our best choice from this point of view is to maxi-
mize the distance between messages.

Consider now condition (iii). If W(e) is even but is allowed to decrease for $e \geq 0$
the case illustrated in Fig. 28 could occur. When one of the message amplitudes is x,
then, by selecting $x \pm e_0$ for two of the other message amplitudes, no weight will be given
to the resulting error. Condition (iii) rules out such a case and appears, therefore, as
a necessary requirement.

We showed in Section VI that, for most error criteria, a message probability density
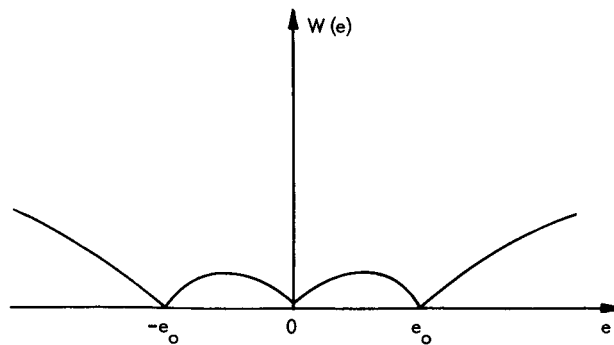made up of 3 impulses will lead to an arbitrarily small W(e) as the maximum message

Fig. 28. Error-weighting function for which condition (iii) is not fulfilled.

76

amplitude L is allowed to go to infinity. Since, as can be easily verified, this result does not occur when the message probability density is made up of 4 distinct impulses, we have, in this limiting case, another indication of the correctness of the result claimed here.

We have carried out computations for a message probability density made up of 4 impulses and $W(e) = e^2$, when the noise probability density has the expression

$$p_n(v) = \begin{cases} 0 & v < -2 \\ \dfrac{1}{4} & -2 \leq v \leq 2 \\ 0 & v > 2. \end{cases}$$

We have $p_m(v)$ as given by Eq. 56 and we require

$$0 \leq x \leq 1, \quad 1 \leq y \leq 2, \quad ax^2 + (1-a) y^2 = 1.$$

We have verified that for any y, $1 \leq y \leq 2$, the mean-square error will be minimum when $x = 0$ or $x = 1$. Since, under the constraint, these values correspond to 3 and 2 message impulses, respectively, we have verified our result for this specific example.

8.4 LOWER BOUNDS ON THE MEAN-SQUARE ERROR

By making use of the expression for the optimum mean-square filter when the message is made up of either 3 or 2 impulses, we shall obtain lower bounds to the mean-square error. Since the 3 impulses reduce to 2 under the average power constraint when $y = M$, we shall study the mean-square error for a fixed M and y varying from M to infinity. For a specific maximum message amplitude L we shall select as lower bound the minimum mean-square error in the range $M \leq y \leq L$.

$$\overline{e^2} = M^2 - \int g^2(x) \, p_{m+n}(x) \, dx.$$

The optimum filter is given by the conditional mean

$$g(x) = \frac{\int v \, p_n(x-v) \, p_m(v) \, dv}{p_{n+m}(x)},$$

and

$$p_m(v) = a \, u(v) + \frac{1-a}{2} \, u(v-y) + \frac{1-a}{2} \, u(v+y). \tag{71}$$

We make use of the power constraint $(1-a) y^2 = M^2$ to obtain

$$\overline{e^2} = M^2 - \frac{M^2}{2} \int \frac{[p_n(x-y)-p_n(x+y)]^2}{2\left[\dfrac{y^2}{M^2} - 1\right] p_n(x) + p_n(x+y) + p_n(x-y)} . \tag{72}$$

Thus we have an expression depending on $p_n(x)$, M, and y which has to be studied for each type of noise. Because of its special importance we carry out the discussion for Gaussian noise.

a. Gaussian Noise

We take $p_n(x) = 1/\sqrt{2\pi} \, \sigma \exp[-x^2/2\sigma^2]$, and, by substitution in expression (72) and some simplification, we can write

$$\overline{e^2} = M^2 - M^2 e^{y^2/2\sigma^2} \int \frac{1}{\sqrt{2\pi}\,\sigma} \frac{\sinh^2 \dfrac{yx}{\sigma^2} \exp\left[-\dfrac{x^2}{2\sigma^2}\right] dx}{\left(\dfrac{y^2}{M^2} - 1\right) \exp\left[\dfrac{y^2}{2\sigma^2}\right] + \cosh \dfrac{yx}{\sigma^2}} . \tag{73}$$

If we let

$$k = \left[\frac{y^2}{M^2} - 1\right] \exp\left[\frac{y^2}{2\sigma^2}\right]$$

and note that we can write $\sinh^2 yx/\sigma^2 = \cosh^2 yx/\sigma^2 - k^2 + k^2 - 1$, then for part of the integrand of (73) we have

$$\frac{\sinh^2 \dfrac{yx}{\sigma^2}}{k + \cosh \dfrac{yx}{\sigma^2}} = \cosh \frac{yx}{\sigma^2} - k + \frac{k^2 - 1}{k + \cosh \dfrac{yx}{\sigma^2}} .$$

Since we have

$$\int \frac{\exp\left[-\dfrac{x^2}{2\sigma^2}\right]}{\sqrt{2\pi}\,\sigma} dx = 1 \qquad \int \frac{\exp\left[-\dfrac{x^2}{2\sigma^2}\right]}{\sqrt{2\pi}\,\sigma} \cosh \frac{yx}{\sigma^2} dx = \exp\left[\frac{y^2}{2\sigma^2}\right],$$

we can write, after some simplification,

$$\frac{\overline{e^2}}{M^2} = \exp\left[-\frac{y^2}{2\sigma^2}\right]\left[k - \frac{k^2 - 1}{\sqrt{2\pi}\,\sigma} \int \frac{\exp\left[-\dfrac{x^2}{2\sigma^2}\right]}{k + \cosh \dfrac{yx}{\sigma^2}} dx\right] . \tag{74}$$

Expression (74) cannot be easily evaluated, but we can conveniently find lower bounds to it. We have to consider separately the cases for $y = M$ and $y \neq M$.

For $y = M$, we have now $k = 0$ and

$$\frac{\overline{e^2}}{M^2} = \frac{\exp\left[-\frac{M^2}{2\sigma^2}\right]}{\sqrt{2\pi}\ \sigma} \int \frac{\exp\left[-\frac{x^2}{2\sigma^2}\right]}{\cosh\frac{Mx}{\sigma^2}}\ dx.$$

If we note that

$$\cosh\frac{Mx}{\sigma^2} \leq e^{Mx/\sigma^2} \qquad \text{for all } x \geq 0,$$

we have

$$\frac{\overline{e^2}}{M^2} \geq \frac{2e^{-M^2/2\sigma^2}}{\sqrt{2\pi}\ \sigma} \int_0^\infty e^{-x^2/2\sigma^2}\ e^{-Mx/\sigma^2}\ dx$$

or

$$\frac{\overline{e^2}}{M^2} \geq \frac{2}{\sqrt{2\pi}} \int_{M/\sigma}^\infty e^{-x^2/2}\ dx = 2\left[1 - \Phi\left(\frac{M}{\sigma}\right)\right],$$

in which we let $\Phi(v) = 1/\sqrt{2\pi} \int_{-\infty}^v e^{-x^2/2}\ dx$.

For $y \neq M$, we shall make 2 approximations in Eq. 74: The first one leads to a very simple expression for the lower bound but gives a rather loose bound; the second one yields a tighter bound.

APPROXIMATION 1: In Eq. 74 we can write

$$\frac{1}{\sqrt{2\pi}\ \sigma} \int \frac{\exp\left[-\frac{x^2}{2\sigma^2}\right]}{k + \cosh\frac{yx}{\sigma^2}} \leq \frac{1}{\sqrt{2\pi}\ \sigma} \int \frac{\exp\left[-\frac{x^2}{2\sigma^2}\right]}{k + 1}\ dx = \frac{1}{k + 1},$$

and we have, therefore, the very simple result

$$\frac{\overline{e^2}}{M^2} \geq \exp\left[-\frac{y^2}{2\sigma^2}\right]. \tag{75}$$

We shall have equality in (75) when $k = 1$ or

$$\frac{y^2}{M^2} - 1 = e^{-y^2/2\sigma^2}. \tag{76}$$

For larger values of y/M inequality (75) is valid. For smaller values of y/M we have,

79

in fact, an upper bound to the expression given in Eq. 74.

For $y \neq M$ the expression of the performance index takes a very simple form, too. Since

$$\eta(k) = \frac{\dfrac{\overline{e^2}}{M^2}(1+k)}{k}$$

with $k = \sigma^2/M^2$, we have here

$$\eta\left(\frac{\sigma^2}{M^2}\right)_{nepers} \geq \frac{y^2}{2M^2}\ln\left(1 + \frac{M^2}{\sigma^2}\right).$$

APPROXIMATION 2: If we note now that $\cosh yx/\sigma^2 \geq 1 + y^2x^2/2\sigma^4$ for all x, we have

$$\frac{\overline{e^2}}{M^2} \geq e^{-y^2/2\sigma^2}\left[k - \frac{k^2-1}{\sqrt{2\pi}\,\sigma}\int \frac{\exp\left[-\dfrac{x^2}{2\sigma^2}\right]}{k+1+\dfrac{y^2x^2}{2\sigma^4}}dx\right].$$

The integral has been tabulated,[19] and we have

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{e^{-x^2/2}}{\beta^2+x^2}dx = \frac{e^{\beta^2/2}}{\beta}\int_{\beta}^{\infty}e^{-x^2/2}dx.$$

If we let $\beta^2 = 2\sigma^2(k+1)/y^2$, we have

$$\frac{\overline{e^2}}{M^2} \geq e^{-y^2/2\sigma^2}\left[k - (k-1)\beta\,e^{\beta^2/2}\int_{\beta}^{\infty}e^{-x^2/2}dx\right].$$

Tabulation of the quantity $e^{\beta^2/2}\int_{\beta}^{\infty}e^{-x^2/2}dx$ is common, and the expression $\alpha = 1 - \beta\,e^{\beta^2/2}\int_{\beta}^{\infty}e^{-x^2/2}dx$ has been tabulated by Sheppard.[20]

In terms of α and k we have

$$\frac{\overline{e^2}}{M^2} \geq e^{-y^2/2\sigma^2}\left[1+\alpha(k-1)\right],$$

which is valid over the same range of L/M as Approximation 1.

EXAMPLE: We take $M/\sigma = 1$. For $y = M$ we have

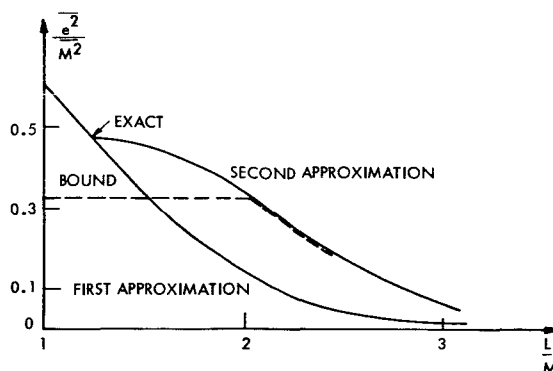$$\frac{\overline{e^2}}{M^2} \geq 2\left[1-\phi(1)\right] = 0.3174.$$

80

Fig. 29. Lower bound on the mean-square error.

For $y \neq M$ the lower bound is valid for $y/M > 1.22$. Approximation 1 gives

$$\frac{\overline{e^2}}{M^2} \geq e^{-y^2/2M^2},$$

and Approximation 2 gives

$$\frac{\overline{e^2}}{M^2} \geq e^{-y^2/2M^2} [1+(k-1)a],$$

in which $k = (y^2/M^2-1)\, e^{y^2/2M^2}$ and a are as defined above. We obtain the graph given in Fig. 29.

8.5 LOWER BOUNDS ON THE ABSOLUTE VALUE OF THE ERROR

In order to obtain lower bounds to the absolute value of the error we have to find the expression for the optimum filter whenever the message probability density consists of either 3 or 2 impulses. We have seen that the optimum filter for the absolute value of the error criterion corresponds to the median of the conditional probability density of the desired output, given the input. The problem of finding the median of a distribution turns out to be quite simple whenever we have a discrete distribution. We have $p_{m/m+n}(v/x) = p_n(x-v)\, p_m(v)/p_{m+n}(x)$ and, since $p_n(v) \neq 0$ for all v, we have $p_{m+n}(x) \neq 0$ for all x and, therefore, $p_{m/m+n}(v/x)$ consists of the impulses for the message probability density $p_m(v)$, multiplied by some weighting factor $p_n(x-v)/p_{m+n}(x)$. Therefore the median occurs at one of the possible values of the message. The only exception is the case for which the impulses to the left of v_{k+1}, the position of one of the impulses, sum to 1/2, and in that case the median v^* is any v such that $v_k \leq v < v_{k+1}$. The filter is not unique, however, and the choice $v^* = v_k$ is still possible. Therefore we can say that, in the case of a quantized message, the output of the optimum filter is quantized at the values of the message, and the noise will only affect the point of transition from one message level to another.[2]

Here we consider 3 impulses

$$p_m(v) = a\, u(v) + \frac{1-a}{2}\, u(v-y) + \frac{1-a}{2} u(v+y)$$

and, therefore, $\overline{W(e)}$ (Eq. 53) is given by

$$\overline{W(e)} = a \int |-g(x)|\, p_n(x)\, dx + \frac{1-a}{2} \int |y-g(x)|\, p_n(x-y)\, dx$$

$$+ \frac{1-a}{2} \int |-y-g(x)|\, p_n(x+y)\, dx$$

and the optimum filter takes the form

$$g(x) = \begin{cases} 0 & \text{for } x \in \Omega_o \\ +y & \text{for } x \in \Omega_1 \\ -y & \text{for } x \in \Omega_2, \end{cases} \tag{77}$$

in which Ω_o, Ω_1, and Ω_2, the regions of x giving each of the possible levels, will depend on the noise characteristic. We can write

$$\overline{W(e)} = ay \int_{\Omega_1 + \Omega_2} p_n(x)\, dx + (1-a)\, y \int_{\Omega_o} p_n(x-y)\, dx + 2(1-a)\, y \int_{\Omega_2} p_n(x-y)\, dx,$$

where we have used the facts that $p_n(x)$ is even and $g(x)$ is odd. Since $(1-a)\, y^2 = M^2$, because of the power constraint we have

$$\overline{W(e)} = \frac{M^2}{y} \left[\left(\frac{y^2}{M^2} - 1 \right) \int_{\Omega_1 + \Omega_2} p_n(x)\, dx + \int_{\Omega_o} p_n(x-y)\, dx + 2 \int_{\Omega_2} p_n(x-y)\, dx \right].$$

We illustrate the determination of regions Ω_o, Ω_1, and Ω_2 for Gaussian noise.

a. Gaussian Noise

The median of $p_{m/m+n}(v/x)$ is zero for x = 0. For increasing x it will jump to the value y for the value x_o such that the magnitude of the impulse at v = +y becomes 1/2. For $x > x_o$ the median keeps the value y for Gaussian noise. Since the magnitude of the impulse at v = +y is $\dfrac{1-a}{2} \dfrac{p_n(x-y)}{p_{n+m}(x)}$, we shall have transition whenever

$$\frac{1-a}{2} p_n(x_o-y) = \frac{1}{2} p_{n+m}(x_o)$$

or

$$(1-a)\, p_n(x_o-y) = a\, p_n(x_o) + \frac{1-a}{2} p_n(x_o-y) + \frac{1-a}{2} p_n(x_o+y).$$

Since

82

$$p_n(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-x^2/2\sigma^2} \qquad \text{and} \qquad (1-a)\, y^2 = M^2,$$

for x_o we find

$$x_o = \frac{\sigma^2}{y}\, \sinh^{-1}\!\left(\frac{y^2}{M^2} - 1\right) e^{y^2/2\sigma^2} = \frac{\sigma^2}{L}\, \sinh^{-1} k,$$

in which k is the parameter defined earlier. Therefore we have $\Omega_o = [-x_o, x_o]$, $\Omega_1 = (x_o, \infty)$, and $\Omega_2 = (-\infty, -x_o)$, and the expression for the absolute value of the error becomes

$$\overline{W(e)} = \frac{M^2}{y}\left[2\left(\frac{y^2}{M^2} - 1\right)\int_{x_o}^{\infty} p_n(x)\,dx + \int_{-x_o}^{+x_o} p_n(x-y)\,dx + 2\int_{-\infty}^{-x_o} p_n(x-y)\,dx\right].$$

Since $p_n(v)$ is Gaussian, we use the function

$$\phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-v^2/2}\,dv,$$

and we have

$$\overline{W(e)} = \frac{M^2}{y}\left\{2\left(\frac{y^2}{M^2} - 1\right)\left[1 - \phi\left(\frac{x_o}{\sigma}\right)\right] + \phi\left(\frac{x_o - y}{\sigma}\right) + \phi\left(\frac{-x_o - y}{\sigma}\right)\right\}.$$

EXAMPLE: We take $M/\sigma = 1$, and we obtain the curve given in Fig. 30. The lower bound is constant up to $L/M = 2$.
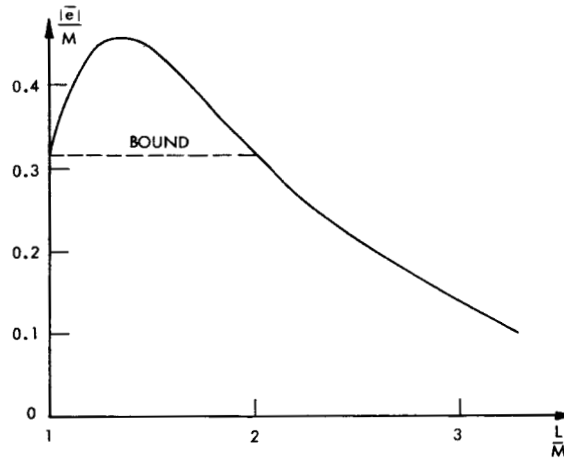

Fig. 30.   Lower bound on the absolute value
of the error.

8.6 LOWER BOUNDS ON THE PROBABILITY THAT THE ABSOLUTE VALUE OF THE ERROR IS LARGER THAN SOME CONSTANT A $(\text{Prob}\{|e| > A\})$

For this criterion of error we select the filter g(x) such that $\int_{-A+g(x)}^{A+g(x)} p_{d/i}(v/x)\,dv$ is as large as possible. If the conditional probability density is discrete and the impulses are farther away than 2A, this integral will yield the magnitude of each of these impulses. Therefore we are led to select for g(x) the most likely value of the desired output, given the input, and the filter gives an output quantized at the values of the desired output. Here, we consider

$$p_{d/i}(v/x) = p_{m/m+n}(v/x) = \frac{p_n(x-v)\,p_m(v)}{p_{m+n}(x)},$$

and the probability density consists of 3 impulses at most, with a minimum distance between impulses which is either 2M or y. Therefore, if A < (the smaller of M, y/2) we have a filter, a resulting average error, and, therefore, lower bounds that are independent of A. We shall confine ourselves to this case, which leads to simple computations. Here, again, the optimum filter g(x) takes the form given by Eq. 77 in which regions Ω_0, Ω_1, and Ω_2 depend on the specific noise. The average weighted error has the form

$$\overline{W(e)} = \iint W[v-g(x)]\,p_n(x-v)\,p_m(v)\,dv\,dx$$

$$= a \int_{\Omega_1+\Omega_2} p_n(x)\,dx + \frac{1-a}{2} \int_{\Omega_0+\Omega_2} p_n(x-y)\,dx + \frac{1-a}{2} \int_{\Omega_0+\Omega_1} p_n(x+y)\,dx.$$

Since $p_n(v)$ is even and we have $(1-a)\,y^2 = M^2$, we can write

$$\overline{W(e)} = \left(1 - \frac{M^2}{y^2}\right) \int_{\Omega_1+\Omega_2} p_n(x)\,dx + \frac{M^2}{y^2} \int_{\Omega_0+\Omega_2} p_n(x-y)\,dx.$$

To find Ω_0, Ω_1, and Ω_2, we have to compare the magnitudes of the impulses of the conditional probability density. We have impulses at v = 0, v = +y, and v = -y and, since the filter is odd, we can limit ourselves to comparing the impulses of v = 0 and v = y. The impulse at v = 0 has a magnitude $b_0 = a\,\dfrac{p_n(x)}{p_{n+m}(x)}$ and, for the impulse at v = y, we have $b_1 = \dfrac{1-a}{2}\,\dfrac{p_n(x-y)}{p_{n+m}(x)}$. For x > 0 the transition from g(x) = 0 to g(x) = y occurs for the input x_0 such that $b_0 = b_1$, that is,

$$a\,p_n(x_0) = \frac{1-a}{2}\,p_n(x_0-y),$$

or, taking into account the power constraint, we have

$$\left(1 - \frac{M^2}{y^2}\right) p_n(x_o) = \frac{M^2}{2y^2} \, p_n(x_o - y).$$

a. Gaussian Noise

Here the transition point x_o is found by the equation

$$\left(1 - \frac{M^2}{y^2}\right) \exp\left[-\frac{x_o^2}{2\sigma^2}\right] = \frac{M^2}{2y^2} \exp\left[-\frac{(x_o - y)^2}{2\sigma^2}\right],$$

which gives

$$x_o = \frac{y}{2} + \frac{\sigma^2}{y} \ln 2\left(\frac{y^2}{M^2} - 1\right). \tag{78}$$

The transition point x_o has to be positive. If Eq. 78 yields $x_o \leq 0$, we have transition at $x_o = 0$ and the filter takes only the values $+y$ and $-y$ and never the value zero. If we let $x_o = 0$ in Eq. 78, we have

$$\frac{y^2}{M^2} - 1 = \frac{1}{2} e^{-y^2/2\sigma^2}. \tag{79}$$

If y^2/M^2 is smaller than the solution of Eq. 79, then the transition from $-y$ to $+y$ occurs at $x = 0$ and we have $\Omega_o = [0]$, $\Omega_1 = (0, \infty)$, $\Omega_2 = (-\infty, 0)$, and

$$\overline{W(e)} = 1 - \frac{M^2}{y^2} + \frac{M^2}{y^2} \int_{-\infty}^{0} p_n(x-y) \, dx$$

or

$$\overline{W(e)} = 1 - \frac{M^2}{y^2} + \frac{M^2}{y^2} \, \phi\left(\frac{-y}{\sigma}\right).$$

When y^2/M^2 is larger than the solution of Eq. 79, we have $\Omega_o = (-x_o, +x_o)$, $\Omega_1 = (x_o, \infty)$, $\Omega_2 = (-\infty, -x_o)$, and the average weighted error becomes

$$\overline{W(e)} = 2\left(1 - \frac{M^2}{y^2}\right) \int_{x_o}^{\infty} p_n(x) \, dx + \frac{M^2}{y^2} \int_{-\infty}^{x_o} p_n(x-y) \, dx$$

or

$$\overline{W(e)} = 2\left(1 - \frac{M^2}{y^2}\right)\left[1 - \phi\left(\frac{x_o}{\sigma}\right)\right] + \frac{M^2}{y^2} \, \phi\left(\frac{x_o - y}{\sigma}\right).$$

EXAMPLE: We take $M = \sigma$. Here the solution of Eq. 79 is $y/M = 1.12$ and there-fore for $y/M > 1.12$ we have $x_o > 0$. The graph of the minimum probability is given
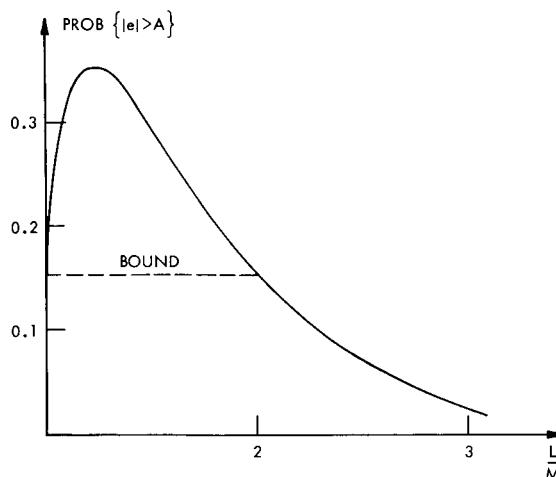

Fig. 31. Lower bound on Prob$\{|e| > A\}$.

in Fig. 31. The lower bound is a constant up to $L/M = 2$. Note that when we use a non-optimum filter giving $g(x) = 0$ for all inputs, the probability of error is the sum of the magnitudes of the impulses at $\pm L$. Because of the power constraint, this is equal to M^2/L^2; therefore the lower bound will be smaller than M^2/L^2 for any noise.

IX. CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

9. 1 CONCLUSIONS

The purpose of this work was to gain some insight into the process of filtering, principally nonlinear filtering. To this end, we asked a set of questions on the behavior of filters which by-passed the difficult analytical problem, in particular, in the case of non mean-square filtering, of determining the optimum filter from the input statistics. In this report, we indicate in which cases linear filtering will be the only useful processing of the input, and our central result is that a crudely quantized message (a message with "nonlinear" characteristics) leads to the best possible separation from the noise for most error criteria. Because of the extremely simple structure of the optimizing message, the problem of lower bounds to the average weighted error can be solved, although it requires knowledge of the optimum filter. With respect to the usefulness of the bounds thus obtained we note that

(i) Since there is a message probability density of the class considered which will give the lower bound, we do have a good or "tight" bound for the constraints used.

(ii) Whether this bound, which corresponds to a quantized message, is a good indication of what to expect, when the message is not quantized, for instance, is a point that requires further investigation. We feel, however, that the bound will be of value as long as the peak-to-average power ratio is not too large. (Note, for instance, that this ratio is equal to three for a flat probability density.) This leads naturally to the discussion of other meaningful message constraints (see section 9. 2b).

More generally, we believe that the investigation of properties of the message which lead to good or poor separation from the noise is a new approach to the problem of filtering. The answers obtained give insight into this difficult problem and provide some indication of the results to be expected. We feel that both insight and motivation are needed before undertaking the lengthy computations for the determination of the optimum filter arising in a complex situation. Similar comments are applicable to the field of nonlinear prediction, and we note that most of the questions raised in this work are relevant in this related field as well.

9. 2 SUGGESTIONS FOR FURTHER RESEARCH

a. Filters with Memory

It is clearly of interest to extend these results to filters with memory and to study, in this more general case, the messages that lead to a poor or a good separation from the noise. Although we shall not undertake here a general discussion of this problem, we shall indicate in a simple case a possible formulation. We shall restrict ourselves to a filter that operates on the input at two different time instants and consider the specific question of upper and lower bounds on the average weighted error for a given filter.

Consider an input signal $x(t)$, $x(t) = m(t) + n(t)$, in which $m(t)$ and $n(t)$ are statistically independent processes. It is desired to estimate the message at times t_1 and

t_2 from the knowledge of the input at the same time instants. The noise statistics are known. Let $\hat{m}_1$, $\hat{m}_2$ be the estimates of $m(t_1)$ and $m(t_2)$, respectively, and let us assume that the filter characteristic is known. We have

$$(\hat{m}_1, \hat{m}_2) = \{g(x_1, x_2), h(x_1, x_2)\},$$

in which $g(x_1, x_2)$ and $h(x_1, x_2)$ are known functions of the input amplitudes x_1 and x_2. Note that $g(x_1, x_2)$ and $h(x_1, x_2)$ are not necessarily unrelated and that, if the input processes are stationary, we should have $g(x_1, x_2) = h(x_2, x_1)$. Let the error-weighting function be a scalar function of two variables such as

$$W(e_1, e_2) = e_1^2 + e_2^2.$$

For the average weighted error we can write

$$\overline{W(e)} = \iiiint W[\lambda_1 - g(x_1, x_2), \lambda_2 - h(x_1, x_2)] \, P_{m_1, m_1+n_1, m_2, m_2+n_2}(\lambda_1, x_1, \lambda_2, x_2)$$

$$\dots \, dx_1 \, d\lambda_1 \, dx_2 \, d\lambda_2,$$

but we have

$$P_{m_1, m_1+n_1, m_2, m_2+n_2}(\lambda_1, x_1, \lambda_2, x_2) = P_{m_1+n_1, m_2+n_2/m_1, m_2}(x_1, x_2/\lambda_1, \lambda_2)$$

$$\times P_{m_1, m_2}(\lambda_1, \lambda_2)$$

$$= P_{n_1, n_2}(x_1 - \lambda_1, x_2 - \lambda_2) \, P_{m_1, m_2}(\lambda_1, \lambda_2).$$

If we let

$$F(\lambda_1, \lambda_2) = \iint W[\lambda_1 - g(x_1, x_2), \lambda_2 - h(x_1, x_2)] \, P_n(x_1 - \lambda_1, x_2 - \lambda_2) \, dx_1 \, dx_2,$$

we have

$$\overline{W(e)} = \iint F(\lambda_1, \lambda_2) \, P_{m_1, m_2}(\lambda_1, \lambda_2) \, d\lambda_1 \, d\lambda_2. \tag{80}$$

It is clear that $F(\lambda_1, \lambda_2)$ is a known function of λ_1 and λ_2 for a given filter. Our problem is to find the specific message probability density $P_{m_1, m_2}(\lambda_1, \lambda_2)$ that will maximize or minimize expression (80), under some message constraints. Of particular interest is the set of constraints

$$\iint \lambda_i \lambda_j \, P_{m_1, m_2}(\lambda_1, \lambda_2) \, d\lambda_1 \, d\lambda_2 = R_{ij}, \tag{81}$$

in which $i, j = 1, 2$ and the correlation matrix of the message R_{ij} is specified. To obtain

useful bounds, we shall most probably need here, also, a constraint of the peak amplitude of the message. For these constraints the problem considered can be expressed as follows.

If we have available the statistical information that is necessary to find the optimum linear filter, and we know, furthermore, the peak amplitude of the message, What is the message most (or least) separable from the noise by a given nonlinear filter (or by the optimum nonlinear filter)? The solution of this problem by a generalization to the case of several variables of the method used in Section VII appears possible but will not be attempted here. We feel that a message quantized at 4 different levels will again be the message giving the extrema of the average weighted error. For the optimum filter, an interesting measure of performance is the ratio of the average weighted error for the best message to the average weighted error achievable by an optimum linear operation on the input. As does the performance index for no-memory filters, this index will indicate the improvement in performance resulting from the nonlinear part of the filter.

b. Other Constraints on the Message

Although in Section VIII we considered specifically an average power constraint on the message, it is clear, from the results of Section VII, that, for all constraints on the message of the form $\int F(v) \, p_m(v) \, dv$ = constant, the optimum message probability density will be made up of 4 impulses, whether or not the optimum filter is used. It is to be expected, however, that in most of the cases for which filtering is to be performed, the message will have a large number of possible amplitudes or possibly a continuous probability distribution. The low number of quantization levels of the optimum message may detract, therefore, from the usefulness, in such cases, of our lower bounds on the average weighted error. It would be desirable to find a constraint that requires that the message probability density be free of impulses, and then find the minimum average weighted error under this constraint. Constraints of the form $\int G[p_m(v)] \, p_m(v) \, dv$ = constant, in which $G[p_m(v)]$ is some appropriate function of $p_m(v)$, have to be investigated. An entropy constraint $\int \log[p_m(v)] \, p_m(v) \, dv$ = constant belongs to this class and may free $p_m(v)$ of impulses, since any impulse will give the value $-\infty$ to the entropy. Because of the physical connotations of entropy and because of its widespread usage in the field of communication, the applications of this specific constraint to the message seems to be a desirable area of future work.

APPENDIX A

DERIVATIVES OF THE NORMALIZED MEAN-SQUARE ERROR OF LINEAR FILTERS WHEN THE NOISE LEVEL k IS THE INDEPENDENT VARIABLE

In order to facilitate the study of derivatives, we define

$$A = \Phi_{nno}(\omega) \qquad B = \Phi_{mm}(\omega).$$

Hence,

$$\beta(k) = \frac{\int \frac{kAB}{kA + B} \, d\omega}{\int B \, d\omega} \, .$$

Since the integrand $\frac{kAB}{kA + B}$ is an integrable function of ω for all values of k, and

$$\frac{\partial}{\partial k} \frac{kAB}{kA + B} = \frac{AB^2}{[kA+B]^2}$$ exists and is a continuous function of ω and k, we can write the

first derivative of $\beta(k)$,

$$\frac{d\beta(k)}{dk} = \frac{\int \frac{AB^2}{[kA+B]^2} \, d\omega}{\int B \, d\omega} \, ,$$

by differentiating with respect to k under the integral sign. A and B are non-negative quantities for all ω; hence $\frac{d\beta(k)}{dk}$ is always positive.

The slope at the origin is

$$\frac{d\beta(k)}{dk}\bigg|_{k=0} = \frac{\int A \, d\omega}{\int B \, d\omega} = \frac{\int \Phi_{nno}(\omega) \, d\omega}{\int \Phi_{mm}(\omega) \, d\omega} = 1.$$

When $k \rightarrow \infty$, the first derivative goes to zero.

The second derivative is

$$\frac{d^2\beta(k)}{dk^2} = \frac{\int \frac{-A^2B^2}{[kA+B]^3} \, d\omega}{\int B \, d\omega} \, .$$

The second derivative is negative for all values of k; hence $\beta(k)$ is a concave function of k.

PROOF OF THE CONCAVITY OF $\Gamma(k)$ VERSUS k

If we use the notation of Appendix A, we can write

$$\Gamma(k) = \frac{\beta(k)}{1-\beta(k)} = \frac{\int \frac{kAB}{kA+B} \, d\omega}{\int \frac{B^2}{kA+B} \, d\omega}$$

$$\frac{d\Gamma(k)}{dk} = \frac{\beta'(k)}{[1-\beta(k)]^2}$$

and

$$\frac{d^2\Gamma(k)}{dk^2} = \frac{[1-\beta(k)]\,\beta''(k) + 2\beta'^2(k)}{[1-\beta(k)]^3}$$

$$= \frac{\int \frac{B^2}{kA+B} \, d\omega \int \frac{-2A^2B^2}{[kA+B]^3} \, d\omega + 2\left|\int \frac{AB^2}{[kA+B]^2} \, d\omega\right|^2}{\left|\int B \, d\omega\right|^2 [1-\beta(k)]^3}.$$

If $\Gamma(k)$ is concave, this second derivative is negative and we need to show that

$$\left|\int \frac{AB^2}{[kA+B]^2} \, d\omega\right|^2 \leq \int \frac{B^2}{kA+B} \, d\omega \int \frac{A^2B^2}{[kA+B]^3} \, d\omega.$$

Let

$$f = \frac{B}{(kA+B)^{1/2}} \qquad g = \frac{AB}{(kA+B)^{3/2}}.$$

We need, therefore,

$$\left|\int fg \, d\omega\right|^2 \leq \int f^2 \, d\omega \int g^2 d\omega,$$

and this is satisfied by the well-known Schwarz inequality.

APPENDIX C

CHARACTERISTIC FUNCTIONS THAT SATISFY $P(t) = [P(ct)]^k e^{j\ell t}$

The relation considered

$$P(t_1) = [P(ct_1)]^k e^{j\ell t_1} \tag{C-1}$$

is interpreted to mean that for any $c > 0$, there is a $k > 0$ and an ℓ such that (C-1) holds. This condition has to be satisfied for all t_1, $-\infty < t_1 < +\infty$. By differentiation, we can write

$$k P'(ct_1) P^{k-1}(ct_1) t_1 e^{j\ell t_1} dc + P^k(ct_1) \ln [P(ct_1)] e^{j\ell t_1} dk$$

$$+ jt_1 P^k(ct_1) d^{j\ell t_1} d\ell = 0,$$

in which $P'(x)$ denotes differentiation with respect to the argument x. Dividing by $P^k(ct_1) e^{j\ell t}$ and letting $ct_1 = t$, we obtain

$$\frac{k}{c} \frac{P'(t)}{P(t)} t \, dc + \ln [P(t)] \, dk + jt \frac{d\ell}{c} = 0,$$

which can be rearranged to give

$$\frac{t P'(t)}{P(t) \ln [P(t)]} + jt \frac{d\ell}{k \, dc} = -\frac{c \, dk}{k \, dc} = A \tag{C-2}$$

in which A is independent of t, k, and c. If k and c are both real, then A is real. We have, therefore, the relations

$$-\frac{1}{A} \frac{dk}{k} = \frac{dc}{c}$$

$$c = D \, k^{-1/a}, \tag{C-3}$$

in which D is a positive constant. We see that there is a one-to-one correspondence between c and k, and that to each positive value of k we can assign a positive value of c.

Instead of solving Eq. C-2 by writing differential equations for the magnitude and the phase of $P(t)$,[21] we shall show that any probability distribution function such that Eq. C-1 is satisfied is a stable distribution function and conversely.

DEFINITION: The distribution function $F(x)$ is called stable if to every $a_1 > 0$, b_1, $a_2 > 0$, b_2 there correspond constants $a > 0$ and b such that the equation $F(a_1 x + b_1)$ * $F(a_2 x + b_2) = F(ax + b)$ holds, where the star denotes convolution. The equation is given in terms of distributions.

For the characteristic function $P(t)$ the defining relation gives

$$P\left[\frac{t}{a_1}\right] P\left[\frac{t}{a_2}\right] = P\left[\frac{t}{a}\right] e^{j\beta t}, \tag{C-4}$$

in which $\beta = b - b_1 - b_2$.

Now we show that Eq. C-1 implies Eq. C-4. Assume that (C-1) is satisfied for $c = a_1$ and $c = a_2$ and gives

$$P(t_1) = [P(a_1 t_1)]^{k_1} e^{j\ell_1 t_1}$$

$$P(t_1) = [P(a_2 t_1)]^{k_2} e^{j\ell_2 t_1},$$

which, by simple changes of variable, become

$$P\left[\frac{t}{a_1}\right] = [P(t)]^{k_1} e^{j\beta_1 t}$$

$$P\left[\frac{t}{a_2}\right] = [P(t)]^{k_2} e^{j\beta_2 t}.$$

Hence,

$$P\left[\frac{t}{a_1}\right] P\left[\frac{t}{a_2}\right] = [P(t)]^{k_1+k_2} e^{j(\beta_1+\beta_2)t}.$$

As we have shown earlier in connection with Eq. C-3, given $k_1 + k_2$, we can find some a and some β such that

$$[P(t)]^{k_1+k_2} e^{j(\beta_1+\beta_2)t} = P\left[\frac{t}{a}\right] e^{j\beta t}$$

Hence, we have Eq. C-4, and hence (C-1) implies (C-4).

To show the converse we use Gnedenko and Kolmogorov's result[11] on the canonical representation of stable distributions.

THEOREM: In order that the distribution function $F(x)$ be stable, it is necessary and sufficient that the logarithm of its characteristic function be represented by the formula

$$\ln P(t) = j\gamma t - d|t|^{\alpha}\{1 + j\delta \, \mathrm{sgnt} \, \omega(t, \alpha)\}, \tag{C-5}$$

where α, δ, γ, and d are constants (γ is any real number, $-1 \leqslant \delta \leqslant 1$, $0 < \alpha \leqslant 2$, $d \geqslant 0$), and

$$\omega(t, \alpha) = \begin{cases} \mathrm{tg} \, \dfrac{\pi}{2} \alpha & \text{if } \alpha \neq 1 \\[2ex] \dfrac{2}{\pi} \ln |t| & \text{if } \alpha = 1. \end{cases}$$

We now shall verify that if the probability distribution is stable, and, hence, Eq. C-5 holds, then Eq. C-1 is satisfied. That is to say, if $\ln P(t)$ is given by Eq. C-5, it is possible to find k and ℓ such that

$$k \ln P(ct) + j\ell t = \ln P(t) \tag{C-6}$$

for any $c > 0$. For stable probability distribution the left-hand member of Eq. C-6 can be written

$$k \ln P(ct) + j\ell t = jk\gamma ct - dk\, c^{a}\, |t|^{a} \left\{ 1 + j\delta\ \text{sgnt}\ \omega(ct, a) \right\} + j\ell t.$$

If $a \neq 1$, then $\omega(ct, a)$ is independent of t. Hence, by taking

$$kc^{a} = 1 \rightarrow k = c^{-a}$$

$$k\gamma c + \ell = \gamma \rightarrow \ell = \gamma[1-kc],$$

we satisfy Eq. C-6. If $a = 1$, we now have

$$\ln P(t) = j\gamma t - d\left\{ |t| + j\delta t \frac{2}{\pi} \ln |t| \right\}$$

$$k \ln P(ct) + j\ell t = j\gamma kct - dk\, c\left\{ |t| + j\delta t \frac{2}{\pi} \ln c\,|t| \right\} + j\ell t.$$

By writing $\ln c(t) = \ln c + \ln (t)$ we have

$$k \ln P(ct) + j\ell t = jt\left[k\gamma c + \ell - dkc\ \delta \frac{2}{\pi} \ln c \right] - dkc\left\{ |t| + j\delta t \frac{2}{\pi} \ln |t| \right\},$$

and we now satisfy Eq. C-6 by taking

$$kc = 1 \rightarrow k = \frac{1}{c}$$

$$k\gamma c + \ell - dkc\delta \frac{2}{\pi} \ln c = \gamma \rightarrow \ell = d\delta \frac{2}{\pi} \ln c.$$

This completes the proof that stable probability distributions that satisfy Eq. C-4 will satisfy Eq. C-1 also.

APPENDIX D

CONDITIONS FOR THE EXISTENCE OF A NONZERO LOWER BOUND ON THE MEAN-SQUARE ERROR

We consider the specific message probability density

$$p_m(x) = (1-2a)\, u(x) + a\, u(x-A) + a\, u(x+A),$$

where $a = M^2/2A^2$.

The optimum filter in the mean-square sense is

$$g_{opt}(x) = \frac{\displaystyle\int v p_m(v)\, p_n(x-v)\, dv}{p_{m+n}(x)}$$

and we have the corresponding mean-square error

$$\overline{e^2} = M^2 - \int g_{opt}^2(x)\, p_{n+m}(x)\, dx$$

By substitution,

$$g(x) = \frac{aA[p_n(x-A)-p_n(x+A)]}{(1-2a)\, p_n(x) + a[p_n(x-A)+p_n(x+A)]}$$

and

$$\overline{e^2} = M^2 - aA^2 \int \frac{[p_n(x-A)-p_n(x+A)]^2\, dx}{(1-2a)\, p_n(x) + a[p_n(x-A)+p_n(x+A)]}\ ,$$

which can be written

$$\overline{e^2} = M^2\left[1 - \int \frac{[p_n(x-A)-p_n(x+A)]^2\, dx}{4\left[\dfrac{A^2}{M^2} - 1\right] p_n(x) + 2[p_n(x-A)+p_n(x+A)]}\right].$$

Let I be the integral within the bracket. We shall show that if the variance of the noise σ_n^2 is finite $I \to 1$ as $A \to \infty$ and, therefore, $\overline{e^2} \to 0$ as $A \to \infty$. The numerator $N(x)$ of the integral I, when expanded, becomes

$$N(x) = p_n^2(x-A) + p_n^2(x+A) - 2p_n(x-A)\, p_n(x+A).$$

Because of the symmetry of the integrand we can write

$$I = \int \frac{[p_n^2(x-A) - p_n(x-A)\,p_n(x+A)]\,dx}{2\left[\dfrac{A^2}{M^2} - 1\right] p_n(x) + p_n(x-A) + p_n(x+A)} \ .$$

We make the change of variable $x - A = y$ and write

$$I = \int \frac{[p_n^2(y) - p_n(y)\,p_n(y+2A)]\,dy}{2\left[\dfrac{A^2}{M^2} - 1\right] p_n(y-A) + p_n(y) + p_n(y+2A)} \ ,$$

or, again,

$$I = \int p_n(x) \ \frac{1 - \dfrac{p_n(y+2A)}{p_n(y)}}{2\left[\dfrac{A^2}{M^2} - 1\right] \dfrac{p_n(y-A)}{p_n(y)} + 1 + \dfrac{p_n(y+2A)}{p_n(y)}} \ dy.$$

Let $M(y)$ be the second part of the integrand. Then, if B is some constant, we write

$$I = \int_{-B}^{B} p_n(y)\,M(y)\,dy + \int_{B}^{\infty} p_n(y)\,M(y)\,dy + \int_{-\infty}^{-B} p_n(y)\,M(y)\,dy.$$

We show now that the first term goes to 1 and the sum of the other two terms becomes arbitrarily small as $B \to \infty$. If we assume that the noise has a finite variance σ_n^2 we can make use of the Tchebycheff inequality

$$\int_{B}^{\infty} p_n(x)\,dx = \int_{-\infty}^{-B} p_n(x)\,dx \le \frac{\sigma_n^2}{2B^2}.$$

For $A > M$ it is easy to see that $-1 \le M(y) \le 1$, and, hence,

$$-\frac{\sigma_n^2}{B^2} \le k \le \frac{\sigma_n^2}{B^2},$$

with

$$k \triangleq \int_{-\infty}^{B} p_n(y)\,M(y)\,dy + \int_{B}^{\infty} p_n(y)\,M(y)\,dy.$$

Now, consider the first term $\int_{-B}^{+B} p_n(y)\,M(y)\,dy$ and take $A \gg B$; then, for $-B \le y \le B$, we can write $\lim\limits_{A \to \infty} M(y) = 1$, if $p_n(y)$ has a finite variance, because we have

$$\lim_{A \to \infty} \frac{p_n(y+2A)}{p_n(y)} = 0$$

$$\lim_{A \to \infty} 2 \left[\frac{A^2}{M^2} - 1 \right] \frac{p_n(y-A)}{p_n(y)} = 0.$$

Hence, by keeping $A \gg B$ in the limiting process, we have

$$\lim_{\substack{A \to \infty \\ B \to \infty}} I = \int_{-B}^{+B} p_n(y) \, M(y) \, dy + k = 1.$$

This, in turn, gives

$$\lim_{A \to \infty} \overline{e^2} = 0.$$

Q. E. D.

APPENDIX E

EXTREMA OF $G = \int_A^B G(x)\, p(x)\, dx$ UNDER THE CONSTRAINT

$$\int_A^B F(x)\, p(x)\, dx = F$$

We would like to find the probability density $p(x)$ that minimizes the integral

$$G = \int_A^B G(x)\, p(x)\, dx \tag{E-1}$$

under the constraint

$$\int_A^B F(x)\, p(x)\, dx = F, \tag{E-2}$$

in which A and B are fixed finite limits, F is a given constant, and $F(x)$ and $G(x)$ are known functions of x.

We try first the classical formulation of the calculus of variations. If we let $p(x) = y^2(x)$ to guarantee positiveness and use the condition $\int_A^B y^2(x)\, dx = 1$, then by the use of the Lagrange multipliers λ and μ we are led to minimize the expression

$$G_1 = \int_A^B [G(x) + \lambda F(x) + \mu]\, y^2(x)\, dx.$$

The corresponding Euler-Lagrange equation gives

$$2y(x)[G(x) + \lambda F(x) + \mu] = 0,$$

which can only be satisfied, in general, by $y(x) = 0$. This implies that there are no solutions satisfying the usual conditions on continuity and differentiability of $y(x)$ which make the Euler-Lagrange formulation valid.

A different approach to the solution of this problem is needed. Let us formulate an auxiliary and related problem. Consider an arbitrary $p(x)$ satisfying the constraint (E-2) and giving some value to the integral of (E-1). For this arbitrary $p(x)$, shown in
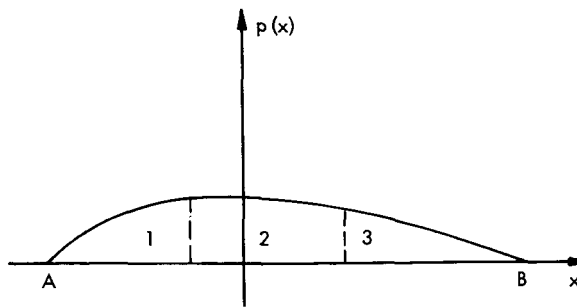

Fig. 32. Division of an arbitrary $p(x)$.

Fig. 32, the regions 1, 2, and 3 are defined by two arbitrary cuts of the x axis between A and B. For this $p(x)$ we shall prove a lemma that establishes that, by proper choice of two of the three regions and normalization, we can form a probability density $p_1(x)$ such that

$$\int_A^B p_1(x)\, dx = 1 \tag{E-3}$$

$$\int_A^B F(x)\, p_1(x)\, dx = F \tag{E-4}$$

$$\int_A^B G(x)\, p_1(x)\, dx \leq \int_A^B G(x)\, p(x)\, dx. \tag{E-5}$$

Similarly, by proper choice of two of the three regions, we can form a probability density $p_2(x)$ such that the two constraints are satisfied and

$$\int_A^B G(x)\, p_2(x)\, dx \geq \int_A^B G(x)\, p(x)\, dx.$$

Before proving this lemma we first define

$$m_1 = \int_{R_1} p(x)\, dx$$

$$F_1 = \int_{R_1} F(x)\, p(x)\, dx / m_1$$

$$G_1 = \int_{R_1} G(x)\, p(x)\, dx / m_1$$

$$m_2 = \int_{R_2} p(x)\, dx$$

$$\vdots$$

in which the indices R_1, R_2, R_3 denote integration over the regions 1, 2, and 3, respectively. Hence, we can write

$$G = m_1 G_1 + m_2 G_2 + m_3 G_3$$

$$F = m_1 F_1 + m_2 F_2 + m_3 F_3$$

$$m_1 + m_2 + m_3 = 1,$$

and the lemma takes the following form.

LEMMA: If m_1, m_2, $m_3 \geq 0$; $m_1 + m_2 + m_3 = 1$; $m_1 F_1 + m_2 F_2 + m_3 F_3 = F$; and $m_1 G_1 + m_2 G_2 + m_3 G_3 = G$, then there exists i and j $[i, j \in (1, 2, 3)]$ and θ_α $[0 \leq \theta_\alpha \leq 1]$ such that

$$\theta_\alpha F_i + (1-\theta_\alpha) F_j = F \qquad \theta_\alpha G_i + (1-\theta_\alpha) G_j \leq G,$$

and there exist k and ℓ $[k, \ell \in (1, 2, 3)]$ and θ_β $[0 \leq \theta_\beta \leq 1]$ such that

$$\theta_\beta F_k + (1-\theta_\beta) F_\ell = F$$

$$\theta_\beta G_k + (1-\theta_\beta) G_\ell \geq G.$$

PROOF: Both parts of the lemma can be proved simultaneously. Assume that $F_1 \leq F$ and F_2, $F_3 \geq F$; then, to fulfill the condition $F_i + (1-\theta) F_j$, we have two possible θ's given by

$$\theta_1 F_1 + (1-\theta_1) F_2 = F$$

$$\theta_2 F_2 + (1-\theta_2) F_3 = F$$

or

$$\theta_1 = \frac{F_2 - F}{F_2 - F_1} \qquad 0 \leq \theta_1 \leq 1$$

$$\theta_2 = \frac{F_3 - F}{F_3 - F_1} \qquad 0 \leq \theta_2 \leq 1.$$

We define

$$M_2 \overset{\Delta}{=} \frac{m_2}{1 - \theta_1} \qquad M_3 \overset{\Delta}{=} \frac{m_3}{1 - \theta_2}$$

and show that we have M_2, $M_3 \geq 0$ and $M_2 + M_3 = 1$. The inequality is clear, since $\theta_1, \theta_2 \leq 1$ and

$$M_2 + M_3 = \frac{m_2}{1 - \theta_1} + \frac{m_3}{1 - \theta_2} = m_2 \left[\frac{F_2 - F_1}{F - F_1} \right] + m_3 \left[\frac{F_3 - F_1}{F - F_1} \right]$$

$$= \frac{m_2 F_2 + m_3 F_3 - F_1(m_2 + m_3)}{F - F_1} = \frac{F - F_1(m_1 + m_2 + m_3)}{F - F_1} = 1.$$

By construction, we have

$$m_2 = M_2(1-\theta_1) \qquad m_3 = M_3(1-\theta_2);$$

hence,

$$m_1 = 1 - (m_2 + m_3) = M_2 \theta_1 + M_3 \theta_2.$$

We now can write

$$G = m_1 G_1 + m_2 G_2 + m_3 G_3 = (M_2 \theta_1 + M_3 \theta_2) G_1 + M_2 (1 - \theta_1) G_2 + M_3 (1 - \theta_2) G_3$$

$$= M_2 [\theta_1 G_1 + (1 - \theta_1) G_2] + M_3 [\theta_2 G_1 + (1 - \theta_2) G_3],$$

and, since $M_2 + M_3 = 1$, we have, necessarily, $\theta_1 G_1 + (1 - \theta_1) G_2 \leq G$ and $\theta_2 G_1 + (1 - \theta_2) G_3 \geq G$ or $\theta_1 G_1 + (1 - \theta_1) G_2 \geq G$ and $\theta_2 G_1 + (1 - \theta_2) G_3 \leq G$.

This establishes the two parts of the lemma for $F_1 \leq F$ and F_2, $F_3 \geq F$. By considering the other case F_1, $F_2 \leq F$ and $F_3 \geq F$ and using a similar reasoning, the proof of this lemma can be completed without difficulty.

We interpret these results in terms of the probability density $p(x)$ by replacing F_1, G_1, and so forth, by their integral expressions. This leads for Eqs. E-4 and E-5 to the expressions

$$\theta_a \frac{\int_{R_i} F(x) p(x) dx}{\int_{R_i} p(x) dx} + (1 - \theta_a) \frac{\int_{R_j} F(x) p(x) dx}{\int_{R_j} p(x) dx} = F$$

$$\theta_a \frac{\int_{R_i} G(x) p(x) dx}{\int_{R_i} p(x) dx} + (1 - \theta_a) \frac{\int_{R_j} G(x) p(x) dx}{\int_{R_j} p(x) dx} \leq G,$$

and the probability density $p_1(x)$ is given by

$$p_1(x) = \theta_a \frac{[p(x)] R_i}{\int_{R_i} p(x) dx} + (1 - \theta_a) \frac{[p(x)] R_j}{\int_{R_j} p(x) dx},$$

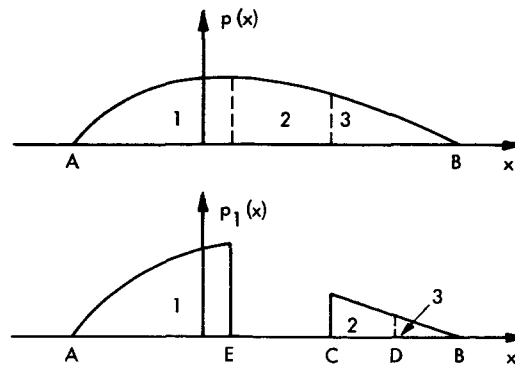in which $[p(x)]_{R_i}$ denotes the portion of the probability density $p(x)$ which belongs to

Fig. 33. Second step in the minimization.

region R_i. Graphically we have, for instance, the case illustrated in Fig. 33 in which region 2 of $p(x)$ has been eliminated.

If we define, in turn, regions 1, 2, and 3 for $p_1(x)$ by the arbitrary cut in D (see Fig. 33) we can apply the lemma to $p_1(x)$ and form a new probability $p_2(x)$ that is non-zero only for two of the three ranges of x defined by AE, CD, DB. This probability $p_2(x)$ will satisfy the constraint and lead to a smaller value of the integral G to be minimized, that is,

$$G_2 = \int_A^B G(x) \, p_2(x) \, dx \leq \int_A^B G(x) \, p_1(x) \, dx = G_1.$$

We see that each successive subdivision and application of the lemma monotonically reduces the base of the probability density which needs to be considered in the minimization. As the number of subdivisions goes to infinity the base of the probability density goes to zero, since the cuts are made arbitrarily. At any intermediary step we have a probability density that consists, at most, of two parts; hence, in the limit the probability density will be made up, at most, of two impulses.

Hence, by successive applications of the lemma we have established that the probability density $p(x)$ which minimizes (or maximizes) Eq. E-1 under the constraint of (E-2) is made up of, at most, two impulses.

Acknowledgment

103

References

1. N. Wiener, *Nonlinear Problems in Random Theory* (The Technology Press of Massachusetts Institute of Technology, Cambridge, Mass., 1958).

2. L. I. Bluestein and R. J. Schwarz, Optimum zero memory filters, IRE Trans., Vol. IT-8, pp. 337-341, October 1962.

3. J. F. Lubbock, The optimization of a class of nonlinear filters, Proc. I. E. E. (London), Part C, Vol. 107, pp. 60-74, March 1960.

4. F. Tung, A Class of Optimum Nonlinear Filters for Quantized Inputs, Technical Report T-2/N, Columbia University, New York, July 1, 1960.

5. A. D. Hause, Nonlinear Least-Squares Filtering and Frequency Modulation, Technical Report 371, Research Laboratory of Electronics, M. I. T., August 25, 1960.

6. See, for instance, Y. W. Lee, *The Statistical Theory of Communication* (John Wiley and Sons, Inc., New York, 1960).

7. G. C. Newton, L. A. Gould, and J. F. Kaiser, *Analytical Design of Linear Feedback Controls* (John Wiley and Sons, New York, 1957).

8. D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions* (John Wiley and Sons, Inc., New York, 1954), Chapter 11.

9. S. Sherman, Non-mean-square error criteria, IRE Trans., Vol. IT-4, pp. 125-126, September 1958.

10. A. V. Balakrishnan, On the characterization of processes for which the optimal mean-square systems are of specified form, IRE Trans., Vol. IT-6, pp. 490-500, September 1960.

11. B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables* (Addison-Wesley Publishing Company, Inc., South Reading, Mass., 1954).

12. *Ibid.*, p. 162.

13. *Ibid.*, p. 181.

14. M. Schetzen, A property of optimum systems, Quarterly Progress Report No. 65, Research Laboratory of Electronics, M. I. T., April 15, 1962, pp. 129-132.

15. B. V. Gnedenko and A. N. Kolmogorov, *op. cit.*, p. 74.

16. F. Tung, Optimal message for zero memory systems with additive Gaussian noise, Proc. IRE 50, 481 (1962).

17. See, for instance, G. Boole, *A Treatise on the Calculus of Finite Differences* (Dover Publications, Inc., New York, 1960), p. 161.

18. See, for instance, G. Birkhoff and G. C. Rota, *Ordinary Differential Equations* (Ginn and Company, Boston, 1962), p. 280.

19. See, for instance, D. Middleton, *An Introduction to Statistical Communication Theory* (McGraw-Hill Book Company, Inc., New York, 1960), Appendix 1, p. 1071.

20. W. F. Sheppard, *The Probability Integral*, B. A. Mathematical Tables Vol. VII (Cambridge University Press, London, 1939).

21. V. R. Algazi, Nonlinear no-memory filters of prescribed form, Quarterly Progress Report No. 66, Research Laboratory of Electronics, M. I. T., July 15, 1962, pp. 201-210.